

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:15:00 ; Search time 53.33 seconds  
(without alignments)  
21.978 Million cell updates/sec

Title: US-09-202-305-13  
Perfect score: 44  
Sequence: 1 IIVTDVIATL 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_unclassified:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	34	77.3	175	13	Q90768 gallus gall
2	34	77.3	203	5	Q9vcv1 drosophila
3	34	77.3	235	2	Q52136 acetobacter
4	33	75.0	178	11	Q9jkn7 marmota mon
5	33	75.0	440	14	Q9j5a5 fowlpox vir
6	33	75.0	990	3	O74202 emericeella
7	33	75.0	990	3	O93862 emericeella
8	32	72.7	81	5	Q9NLR8 leishmania
9	32	72.7	214	2	Q9KY15 streptomyce
10	32	72.7	267	2	Q9PJT0 chlamydia m
11	32	72.7	309	2	Q9KRS5 vibrio chol
12	32	72.7	319	5	Q9XV01 caenorhabdi
13	32	72.7	423	2	Q49939 mycobacteri
14	32	72.7	898	3	O94195 picilia angu
15	31	70.5	58	2	O9ZAT1 streptococ
16	31	70.5	66	2	O32951 mycobacteri
17	31	70.5	227	5	O61790 caenorhabdi
18	31	70.5	331	5	Q9XV83 caenorhabdi
19	31	70.5	342	5	O02125 caenorhabdi

20	31	70.5	350	5	Q19556	Q19556 caenorhabdi
21	31	70.5	378	2	Q9KR60	Q9kr60 vibrio chol
22	31	70.5	393	5	Q46077	Q46077 drosophila
23	31	70.5	397	5	Q9W504	Q9w504 drosophila
24	31	70.5	406	10	Q9SIE1	Q9siel arabidopsis
25	31	70.5	430	2	Q9REU1	Q9reu1 streptomyce
26	31	70.5	430	2	Q9JNS6	Q9jns6 streptomyce
27	31	70.5	669	11	O35540	O35540 mus musculu
28	31	70.5	1095	4	O43156	O43156 homo sapien
29	31	70.5	1192	5	O76373	O76373 caenorhabdi
30	31	70.5	2535	10	O81018	O81018 arabidopsis
31	30	68.2	122	2	Q9Z5J2	Q9z5j2 mycobacteri
32	30	68.2	141	10	Q9ZK43	Q9zk43 medicago tr
33	30	68.2	173	2	Q9X1A4	Q9x1a4 thermotoga
34	30	68.2	192	14	Q89367	Q89367 paramecium
35	30	68.2	205	5	O23193	O23193 caenorhabdi
36	30	68.2	210	2	O9RU32	O9ru32 deinococcus
37	30	68.2	218	1	O59503	O59503 pyrococcus
38	30	68.2	226	1	O58257	O58257 pyrococcus
39	30	68.2	309	2	Q9ZD99	Q9zd99 rickettsia
40	30	68.2	319	5	O45639	O45639 caenorhabdi
41	30	68.2	324	5	O45615	O45615 caenorhabdi
42	30	68.2	388	2	Q9XDW8	Q9xdw8 streptococ
43	30	68.2	417	5	O9VR81	O9vr81 drosophila
44	30	68.2	426	2	Q9XDF7	Q9xdf7 streptomyce
45	30	68.2	495	2	O83577	O83577 treponema p

ALIGNMENTS

RESULT 1

Q90768 ID Q90768 PRELIMINARY; PRT; 175 AA.

AC Q90768; 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE CD3 GLYCOPROTEIN.

GN CD3G/D.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;

OC Gallus.

OC NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91172844; PubMed=1826056;

RA Bernot A., Auffray C.;

RT "Primary structure and ontogeny of an avian CD3 transcript."

RL Proc. Natl. Acad. Sci. U.S.A. 88:2550-2554(1991).

RN [2]

RP SEQUENCE FROM N.A.

RA Goebel T.W., Dangy J.P.;

RT "Evidence for a stepwise evolution of the CD3 family."

RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; M59925; AAA48660.1; -.

DR EMBL; AJ250458; CAB62062.1; -.

DR INTERPRO; IPR003110; -.

DR PFAM; PF02189; ITAM; 1.

DR PRODOM; PD004593; -; 1.

SQ SEQUENCE 175 AA; 19366 MW; 7429EBB88320C2E0 CRC64;

Query Match 77.3%; Score 34; DB 13; Length 175;  
Best Local Similarity 60.0%; Pred. No. 31;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IIVTDVIATL 10  
| | | | |  
Db 112 IIVADVAVT 121

```

RESULT 2
ID Q9VCV1 PRELIMINARY; PRT; 203 AA.
AC Q9VCV1;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE CG6982 PROTEIN.
DE CG6982.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananadis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003740; AAF56054.1; -.
DR FLYBASE; FBgn0039014; CG6982.
SQ SEQUENCE 203 AA; 22693 MW; F346559595D1055C9 CRC64;

```

```

Query Match 77.3%; Score 34; DB 5; Length 203;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 IIVTDVIATL 10
Db :|||||:
93 LIIVTDVIATV 101

```

```

RESULT 3
Q52136 PRELIMINARY; PRT; 235 AA.
ID Q52136

```

```

AC Q52136;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE ORF1.
DE ORF1.
OS Acetobacter sp.
OC Plasmid PAH4.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Acetobacter.
OX NCBI_TaxID=440;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BPR2001.
RX MEDLINE=95072320; PubMed=7765516;
RA Tonouchi N., Tsuchida T., Yoshinaga F., Horinouchi S., Beppu T.;
RT "A host-vector system for a cellulose-producing Acetobacter strain."
RL Biosci. Biotechnol. Biochem. 58:1899-1901(1994).
DR EMBL; D30784; BAA06447.1; -.
DR INTERPRO; IPR001395; -.
DR PROSITE; PS00063; ALDO-KETO-REDUCTASE_3; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 235 AA; 26774 MW; 0EF6B34AD904BE9B CRC64;

```

```

Query Match 77.3%; Score 34; DB 2; Length 235;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 IIVTDVIAT 9
Db :||||:
80 IIVTDVIVAT 88

```

```

RESULT 4
Q9JKN7 PRELIMINARY; PRT; 178 AA.
ID Q9JKN7
AC Q9JKN7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE CD3 EPSTILON CHAIN (FRAGMENT).
OS Marmota monax (Woodchuck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;
OC Marmota.
OX NCBI_TaxID=9995;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20240031; PubMed=10775584;
RA Michalak T.I., Hodgson P.D., Churchill N.D.;
RT "Posttranscriptional inhibition of class I major histocompatibility
complex presentation on hepatocytes and lymphoid cells in chronic
J. Virol. 74:4483-4494(2000).
DR EMBL; AF232727; AAF68959.1; -.
FT NON_TER 178
FT NON_TER 178
SQ SEQUENCE 178 AA; 20019 MW; 45A38B352BB3CE7F CRC64;

```

```

Query Match 75.0%; Score 33; DB 11; Length 178;
Best Local Similarity 60.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 IIVTDVIATL 10
Db :||||:
116 VIVVDIIVTL 125

```

```

RESULT 5
Q9J5A5 PRELIMINARY; PRT; 440 AA.
ID Q9J5A5
AC Q9J5A5;

```

DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE ORF FV117.  
 GN FV117.  
 OS Fowlpox virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Avipoxvirus.  
 OX NCBI\_TaxID=10261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20193820; PubMed=10729156;  
 RA Alfonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;  
 RT "The genome of Fowlpox virus."  
 RL J. Virol. 74:3815-3831(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Alfonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF198100; AAF44461.1;  
 SQ SEQUENCE 440 AA; 51486 MW; C330B8A480D09081 CRC64;

Query Match 75.0%; Score 33; DB 14; Length 440;  
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 IIVTDVIATL 10  
 I:||||:|  
 Db 147 IITDVLASL 155

RESULT 6  
 O74202 PRELIMINARY; PRT; 990 AA.  
 ID O74202  
 AC O74202;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE P-ATPASE.  
 GN PMAA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=5072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Reoyo E., Espeso E., Penalva M.A., Suarez T.;  
 RL "A. nidulans putative plasma membrane H<sup>+</sup>-ATPase."  
 RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF036763; AAC27991.1;  
 DR INTERPRO; IPR000695;  
 DR INTERPRO; IPR001066;  
 DR INTERPRO; IPR001757;  
 DR PFAM; PF00122; EI-E2\_ATPase; 2.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PRINTS; PR00120; HATPASE.  
 DR PROSITE; PS00154; ATPASE\_E1\_E2; UNKNOWN\_1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
 SQ SEQUENCE 990 AA; 108687 MW; E8B2F04ED6459B4C CRC64;

Query Match 75.0%; Score 33; DB 3; Length 990;  
 Best Local Similarity 70.0%; Pred. No. 3.2e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 IIVTDVIATL 10  
 I:||||:|  
 Db 870 IFVVDVLATL 879

RESULT 7  
 O93862 PRELIMINARY; PRT; 990 AA.  
 ID O93862

AC O93862;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE PLASMA MEMBRANE H<sup>+</sup>ATPASE.  
 GN PMAA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=5072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GLASGOW COLLECTION (PABA AL);  
 RA Abdallah B.M., Gorfer M., Strauss J., Kubicek C.P.;  
 RT "Cloning and Characterization of the Aspergillus nidulans Plasma Membrane H<sup>+</sup>-ATPase."  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF043332; AAD11605.1;  
 DR INTERPRO; IPR000695;  
 DR INTERPRO; IPR001066;  
 DR INTERPRO; IPR001757;  
 DR PFAM; PF00122; EI-E2\_ATPase; 2.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PRINTS; PR00120; HATPASE.  
 DR PROSITE; PS00154; ATPASE\_E1\_E2; UNKNOWN\_1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
 SQ SEQUENCE 990 AA; 108754 MW; FFCDD28DE12DD533D CRC64;

Query Match 75.0%; Score 33; DB 3; Length 990;  
 Best Local Similarity 70.0%; Pred. No. 3.2e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 IIVTDVIATL 10  
 I:||||:|  
 Db 870 IFVVDVLATL 879

RESULT 8  
 O9NLR8 PRELIMINARY; PRT; 81 AA.  
 ID O9NLR8  
 AC O9NLR8;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HYPOTHETICAL 8.1 KDA PROTEIN (FRAGMENT).  
 GN LM26.374.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL160493; CAB97992.1;  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 81 AA; 8093 MW; C2D2838FA3B2A91F CRC64;

Query Match 72.7%; Score 32; DB 5; Length 81;  
 Best Local Similarity 70.0%; Pred. No. 35;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 IIVTDVIATL 10  
 I:||||:|  
 Db 19 INVTDVATL 28

RESULT 9  
 O9KY15 PRELIMINARY; PRT; 214 AA.  
 ID O9KY15  
 AC O9KY15;

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE GTP CYCLOHYDROLASE II.  
 GN RIBA.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Oliver K., Harris D.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Redenbach M., Kleser H.M., Denapaita D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RA "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL356595; CAB92253.1; -;  
 KW Hydrolase.  
 SQ SEQUENCE 214 AA; 23429 MW; 318BAEB45289963C CRC64;

Query Match 72.7%; Score 32; DB 2; Length 214;  
 Best Local Similarity 44.4%; Pred. No. 99;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IIIVTDVIAT 9  
 Db 181 VVISDVAT 189

RESULT 10  
 Q9PJTO PRELIMINARY; PRT; 267 AA.  
 AC Q9PJTO  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE TRNA PSEUDOURIDINE SYNTHASE A.  
 GN TC0748.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MOPN / NIGG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumonae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,  
 RA Hickey E.K., Peterson J., Umayam L.A., Utterback T., Berry K.,  
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
 RA Dodson R., Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,  
 RA Salzberg S.L., Eisen J., Fraser C.M.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE002343; AAF39555.1; -;  
 DR TIGR; TC0748; -;  
 DR INTERPRO; IPR001406; -;  
 DR PFAM; PF01416; PseudoU\_synth\_1;  
 SQ SEQUENCE 267 AA; 30413 MW; CBEDB3B568EDBB7 CRC64;

Query Match 72.7%; Score 32; DB 2; Length 267;  
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 IIIVTDVIAT 9  
 Db 91 IVIVTDVIAT 99

RESULT 11  
 Q9KRS5 PRELIMINARY; PRT; 309 AA.  
 AC Q9KRS5  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE TRANSCRIPTIONAL REGULATOR, LYSR FAMILY.  
 GN VC1561.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL; AE004233; AAF94715.1; -;  
 DR TIGR; VC1561; -;  
 SQ SEQUENCE 309 AA; 35940 MW; 89810228C8D6FBE4 CRC64;

Query Match 72.7%; Score 32; DB 2; Length 309;  
 Best Local Similarity 70.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 IIIVTDVIATL 10  
 Db 240 LINTDLIATL 249

RESULT 12  
 Q9XV01 PRELIMINARY; PRT; 319 AA.  
 AC Q9XV01  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE F49H6.10 PROTEIN.  
 GN F49H6.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Lloyd C.;



RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]

RX MEDLINE-94150718; PubMed-7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,  
Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL; Z81545; CAB04443.1;  
SQ SEQUENCE 319 AA; 37742 MW; 7F3D39058EAE5CFF CRC64;

Query Match 72.7%; Score 32; DB 5; Length 319;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIIVTDVIAT 9  
:|:|||||  
Db 259 MIITDVIT 267

RESULT 13

Q49939 PRELIMINARY; PRT; 423 AA.  
AC Q49939;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE HYPOTHETICAL 45.6 KDA PROTEIN UL518A.  
GN UL518A OR L518\_F2\_48.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Robison K., Smith D.R.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: STRONG, TO M. BOVIS HYPOTHETICAL 38.1 KDA PROTEIN IN MAS  
3'REGION (002279).  
CC EMBL; U00023; AAA17363.1;  
DR INTERPRO; IPR001899;  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 423 AA; 45594 MW; FED1483E87F4542E CRC64;

Query Match 72.7%; Score 32; DB 2; Length 423;  
Best Local Similarity 55.6%; Pred. No. 2.1e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 IIVTDVIATL 10  
:|:|||||  
Db 304 LATDIVATL 312

RESULT 14

O94195 PRELIMINARY; PRT; 898 AA.  
AC O94195;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE PLASMA MEMBRANE H+-ATPASE.  
GN PMAL

OS Pichia angusta (Yeast) (Hansenula polymorpha).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Pichia.  
OX NCBI\_TaxID=4905;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A16;  
RA Cox H., Mannazzu I., Evans L., Sudbery P.E.;  
RL "Sequence of the Hansenula polymorpha PMAL gene.";  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF109913; AAD19960.1;  
DR INTERPRO; IPR000695;  
DR INTERPRO; IPR001757;  
DR PFAM; PF00122; EI-E2\_ATPase; 1.  
DR PRINTS; PR00119; CATATPASE.  
DR PRINTS; PR00120; HATPASE.  
SQ SEQUENCE 898 AA; 98098 MW; 0A07D431415128A7 CRC64;

Query Match 72.7%; Score 32; DB 3; Length 898;  
Best Local Similarity 40.0%; Pred. No. 4.6e+02;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIIVTDVIATL 10  
:|:|||||  
Db 809 VLIIVDIIATM 818

RESULT 15

Q9ZAT1 PRELIMINARY; PRT; 58 AA.  
AC Q9ZAT1;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
DE PUTATIVE YIEG PROTEIN (FRAGMENT).  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS-5 KURAMITSU;  
RA Peruzzi F., Piggot P.J., Daneo-Moore L.;  
RT "Development of an integrative, lacZ transcriptional-fusion plasmid  
vector for Streptococcus mutans and its use to identify sugar  
regulated genes.";  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U78603; AAD00284.1;  
FT NON\_TER 58  
SQ SEQUENCE 58 AA; 5894 MW; 38347D11A6523928 CRC64;

Query Match 70.5%; Score 31; DB 2; Length 58;  
Best Local Similarity 80.0%; Pred. No. 39;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IIIVTDVIATL 10  
:|:|||||  
Db 45 IIIVTVIALL 54

Search completed: May 10, 2001, 10:15:02  
Job time: 261 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:10:34 ; Search time 51.82 Seconds  
(without alignments)  
11.031 Million cell updates/sec

Title: US-09-202-305-13  
Perfect score: 44  
Sequence: 1 IIVTDVIATL 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	10	19	W42364
2	44	100.0	171	16	R78669
3	44	100.0	171	17	R89446
4	34	77.3	171	13	R27279
5	34	77.3	171	17	W02216
6	34	77.3	171	20	W83143
7	31	70.5	275	13	R22130
8	31	70.5	308	21	B41868
9	31	70.5	332	21	Y70723
10	31	70.5	397	21	B26448
11	31	70.5	398	21	B20915
					TCR-CD3 (epsilon/delta)
					CD3 delta receptor
					CD3 delta receptor
					T cell receptor pr
					T cell receptor CD
					CD3 delta receptor
					Sequence of ovine
					Human ORFX ORF1632
					Klebsiella oxytoca
					Drosophila melanog
					Drosophila odorant

12	31	70.5	421	21	G30315	Arabidopsis thalia
13	31	70.5	428	21	G30314	Arabidopsis thalia
14	31	70.5	475	21	G30313	Arabidopsis thalia
15	31	70.5	518	21	B53322	Human colon cancer
16	31	70.5	669	19	W37483	Mouse liver cancer
17	31	70.5	671	21	Y99426	Human PRO1604 (UNQ
18	30	68.2	91	18	W01826	Component of bioc
19	30	68.2	91	20	W93819	Angiotropin relate
20	30	68.2	132	21	G14956	Arabidopsis thalia
21	30	68.2	137	21	G14955	Arabidopsis thalia
22	30	68.2	154	21	G07713	Arabidopsis thalia
23	30	68.2	154	21	G12848	Arabidopsis thalia
24	30	68.2	154	21	G47738	Arabidopsis thalia
25	30	68.2	183	14	R30826	Sequence of adenin
26	30	68.2	183	21	G07712	Arabidopsis thalia
27	30	68.2	183	21	G12847	Arabidopsis thalia
28	30	68.2	183	21	G47737	Arabidopsis thalia
29	30	68.2	210	21	G07711	Arabidopsis thalia
30	30	68.2	243	21	G12846	Arabidopsis thalia
31	30	68.2	243	21	G47736	Arabidopsis thalia
32	30	68.2	293	21	G14113	Arabidopsis thalia
33	30	68.2	304	21	G14112	Arabidopsis thalia
34	30	68.2	309	21	G14111	Arabidopsis thalia
35	30	68.2	358	20	Y28280	Human G-protein co
36	30	68.2	358	21	Y45039	Human G-protein co
37	30	68.2	358	21	Y44365	Enterococcus faeca
38	29	65.9	261	20	Y00079	Enterococcus faeca
39	29	65.9	284	20	Y00078	African green monk
40	29	65.9	342	20	W97784	Pig-tailed macaque
41	29	65.9	342	20	W97785	Human prostate can
42	29	65.9	356	21	B56748	Streptococcus pneu
43	29	65.9	399	21	Y81561	Drosophila kinesin
44	29	65.9	411	19	W72745	Drosophila kinesin
45	29	65.9	441	19	W72744	Drosophila kinesin

## ALIGNMENTS

RESULT 1  
W42364  
ID W42364 standard; peptide; 10 AA.  
XX  
AC W42364;  
XX  
DT 11-JUN-1998 (first entry)  
XX  
DE TCR-CD3 (epsilon/delta) chain derived peptide 1.  
XX  
KW Peptide; TCR-CD3 (epsilon/delta) chain; T-cell receptor; inflammation;  
KW autoimmune disease; inflammatory bowel disease; psoriasis; infection;  
KW acquired immune deficiency syndrome; allergy.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO9747644-A1.  
XX  
PD 18-DEC-1997.  
XX  
PF 11-JUN-1997; 97WO-AU00367.  
XX  
PR 11-JUN-1996; 96AU-0000394.  
PR 11-JUN-1996; 96AU-0000389.  
PR 11-JUN-1996; 96AU-0000390.  
PR 11-JUN-1996; 96AU-0000391.  
PR 11-JUN-1996; 96AU-0000392.  
PR 11-JUN-1996; 96AU-0000393.  
XX  
(NSYD-) NORTHERN SYDNEY AREA HEALTH SERVICE.  
XX  
Manolios N;  
XX

DR WPI; 1998-052238/05.  
 XX  
 PT New peptide(s) that inhibit the T cell receptor - used to treat  
 PT inflammation, auto-immune disease, allergy etc. and to deliver  
 PT conjugated therapeutic agents to cells  
 XX  
 PS Claim 7; Page 32; 58pp; English.  
 XX  
 CC The present peptide sequence is derived from the T-cell receptor  
 CC (TCR)-CD3 (epsilon/delta) chain. This peptide and others  
 CC (see W42357-W42371) act to inhibit TCR function probably by interfering  
 CC with TCR assembly. They are claimed to be useful in treating disorders  
 CC in which T cells are involved or recruited, e.g. allergy, autoimmune  
 CC diseases, inflammatory bowel disease, psoriasis, infections (including  
 CC acquired immune deficiency syndrome) etc. They are also claimed to be  
 CC able to deliver conjugated therapeutic agents to cells.  
 XX  
 SQ Sequence 10 AA;  
 Query Match 100.0%; Score 44; DB 19; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0065;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IIVTDVIATL 10  
 DB 1 IIVTDVIATL 10  
 RESULT 2  
 R78669  
 ID R78669 standard; Protein; 171 AA.  
 XX  
 AC R78669;  
 XX  
 DT 11-APR-1996 (first entry)  
 XX  
 DE CD3 delta receptor protein.  
 XX  
 KW Chimeric receptor; CD4; T-cell receptor; HIV; cytotoxicity;  
 KW human immunodeficiency virus; adoptive immunotherapy;  
 KW CD3 delta receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 132..171  
 FT /note= "amino acids 132-171 constitute a region  
 FT sufficient for cytolytic signal  
 FT transduction"  
 XX  
 PN W09521528-A1.  
 XX  
 PD 17-AUG-1995.  
 XX  
 PF 12-JAN-1995; 95WO-US00454.  
 XX  
 PR 02-AUG-1994; 94US-0284391.  
 PR 14-FEB-1994; 94US-0195395.  
 XX  
 PA (GEO) GEN HOSPITAL CORP.  
 XX  
 PI Banapour B, Kolanus W, Romeo C, Seed B;  
 XX  
 DR WPI; 1995-292893/38.  
 XX  
 PT Target cytotoxicity of HIV-infected cells - by chimeric CD4  
 PT receptor-bearing cells  
 XX  
 PS Example 11; Fig 16; 118pp; English.  
 XX  
 CC Intracellular and transmembrane signal transducing domains are  
 CC derived from the T-cell receptor proteins CD3 delta (R78669) and

CC T3 gamma (R78670) and from the B-cell receptor proteins mbl  
 CC (R78671) and B29 (R78672). Each may be linked to the  
 CC extracellular domain of CD4 to obtain a chimeric receptor useful  
 CC for targeted cytotoxicity of HIV-infected cells.  
 XX  
 SQ Sequence 171 AA;  
 Query Match 100.0%; Score 44; DB 16; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 0.17;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IIVTDVIATL 10  
 DB 107 IIVTDVIATL 116  
 RESULT 3  
 R89446  
 ID R89446 standard; Protein; 171 AA.  
 XX  
 AC R89446;  
 XX  
 DT 26-SEP-1996 (first entry)  
 XX  
 DE CD3 delta receptor.  
 XX  
 KW CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;  
 KW human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;  
 KW dendritic cell; therapy; mammal; infection; CD3 delta; T3 gamma; mbl;  
 KW B29.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 132..171  
 FT /note= "cytolytic signal transducing domain"  
 XX  
 PN W09603883-A1.  
 XX  
 PD 15-FEB-1996.  
 XX  
 PF 26-JUL-1995; 95WO-US09468.  
 XX  
 PR 24-FEB-1995; 95US-0394388.  
 PR 02-AUG-1994; 94US-0284391.  
 XX  
 PA (GEO) GEN HOSPITAL CORP.  
 XX  
 PI Banapour B, Kolanus W, Romeo C, Seed B;  
 XX  
 DR WPI; 1996-129034/13.  
 XX  
 PT Membrane-bound chimeric receptor comprising extracellular portion  
 PT including CD4 fragment - cells expressing receptor can be used for  
 PT treatment of HIV infection  
 XX  
 PS Example 12; Fig 16; 134pp; English.  
 XX  
 CC R89446-R89449 represent intracellular and transmembrane signal  
 CC transducing domains that can be used in the membrane bound proteinaceous  
 CC chimeric receptor of the invention. This sequence represents the CD3  
 CC delta receptor protein. Alternatively the transmembrane region of the  
 CC chimeric receptor contains a portion of the CD7, CD5 or CD34  
 CC transmembrane domains. The extracellular portion of the receptor can  
 CC also be separated from the intracellular domain by the hinge, CH2 and CH3  
 CC domains of human IgG1. The extracellular portion of the chimeric  
 CC receptor contains a fragment of CD4 (amino acids 1-394 or 1-200 of the  
 CC CD4 sequence) which specifically recognises and binds HIV-infected cells,  
 CC but does not mediate HIV infection. The extracellular domain of the  
 CC receptor is separated from the cell membrane by 48 or 72 angstroms, or by  
 CC one or more proteinaceous alpha-helices. The cells expressing the  
 CC receptor are preferably T cells, B cells, neutrophils, or dendritic

CC cells. The therapeutic cells expressing the chimeric receptor are  
 CC administered to a mammal to treat HIV infection.

XX Sequence 171 AA;  
 CC  
 Query Match 100.0%; Score 44; DB 17; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 0.17; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IIVTDVIATL 10  
 Db 107 iivtdviall 116

RESULT 4  
 R27279  
 ID R27279 standard; Protein; 171 AA.  
 AC R27279;  
 XX  
 DT 31-JUL-1995 (first entry)  
 XX  
 DE T cell receptor protein, CD3 delta.  
 XX  
 KW Fusion protein; CD4; extracellular domain; zeta; eta; gamma; primer;  
 KW polymerase chain reaction; PCR; amplify; human; IgG1; receptor; chimera;  
 KW heavy; light; chain; tonsil; promoter; bicistronic mRNA; grp78;  
 KW 5' untranslated region; 78 kb glucose related protein; COS cell;  
 KW immune system; pathogen; tumour cell; HIV; cytotoxic T lymphocyte.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 132..171  
 FT /note= "Preferred cytolytic signal transducing  
 FT portion"

XX W09215322-A.  
 XX 17-SEP-1992.  
 XX 06-MAR-1992; 92WO-US01785.  
 XX 07-MAR-1991; 91US-0665961.  
 XX 06-MAR-1992; 92WO-US01785.  
 XX (GEO) GEN HOSPITAL CORP.  
 XX Kolanus W, Romeo C, Seed B;  
 XX WPI; 1992-331474/40.  
 XX  
 PT Therapeutic cells expressing chimeric receptors - directing  
 PT cellular response to an infective agent, useful in treating  
 PT HIV-1, AIDS Pneumocystis carinii infections etc.  
 XX  
 PS Example 10; Fig 16; 114pp; English.

CC The sequences given in R27279-82 represent the intracellular and  
 CC transmembrane signal transducing domains derived from the T cell  
 CC receptor protein CD3 delta and T3 gamma, and the B cell receptor  
 CC proteins, mb1 and B29. Chimeric receptor proteins may be prepared  
 CC by the method of the invention which contain at least the portion of  
 CC the sequence which is sufficient for cytolytic signal transduction.  
 CC immune system cells "armed" with a plasmid encoding a chimera such  
 CC as this, would respond to the presence of the pathogen appropriate  
 CC to their lineage or with tumour cells and immune response could be  
 CC beneficially elevated. This method may be used to direct cellular  
 CC response to an HIV infected cell by administering to a patient an  
 CC effective amount of cytotoxic T lymphocytes which are capable of  
 CC specifically recognising and lysing cells infected with HIV, and can  
 CC therefore be used in the treatment of HIV infected individuals.

XX Sequence 171 AA;  
 SQ  
 Query Match 77.3%; Score 34; DB 13; Length 171;  
 Best Local Similarity 80.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 IIVTDVIATL 10  
 Db 107 iivtdviall 116

RESULT 5  
 W02216  
 ID W02216 standard; Protein; 171 AA.  
 XX  
 AC W02216;  
 XX  
 DT 11-NOV-1996 (first entry)  
 XX  
 DE T cell receptor CD3 delta protein.  
 XX  
 KW Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;  
 KW human immunodeficiency virus type 1; HIV-1; AIDS; therapy;  
 KW T-cell receptor; CD3 delta; cytotoxic T lymphocyte; CTL.  
 XX  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Domain 132..171  
 FT /label= Cytolytic\_signal\_transduction\_domain  
 FT  
 PN W09625953-A1.  
 XX  
 PD 29-AUG-1996.  
 XX  
 PF 25-JAN-1996; 96WO-US01056.  
 XX  
 PR 24-FEB-1995; 95US-0394176.  
 XX  
 PA (GEO) GEN HOSPITAL CORP.  
 XX  
 PI Kolanus W, Romeo C, Seed B;  
 XX  
 DR WPI; 1996-402134/40.  
 XX  
 PT Direction of cellular immune response using therapeutic cell  
 PT expressing 2 chimaeric receptors - comprising region binding to  
 PT target cell and region that signals target cell destruction, or CD28  
 PT region, partic. for eliminating HIV-infected cells

XX Example 10; Fig 16; 120pp; English.  
 XX  
 CC Novel chimaeric receptors may contain intracellular and  
 CC transmembrane signal transducing domains derived from the T cell  
 CC receptor proteins CD3 delta (W02216) or T3 gamma (W02217), or the  
 CC B cell receptor proteins mb1 (W02218) or B29 (W02219), and an  
 CC extracellular domain that allows specific recognition of, and  
 CC binding to, a target cell or target infectious agent. Chimaeric  
 CC receptors (see also W02213-15) can be used to redirect the  
 CC cellular immunity of a mammal.

XX Sequence 171 AA;  
 SQ  
 Query Match 77.3%; Score 34; DB 17; Length 171;  
 Best Local Similarity 80.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 IIVTDVIATL 10  
 Db 107 iivtdviall 116

RESULT 6  
 ID W83143 standard; Protein; 171 AA.  
 XX AC W83143;  
 XX DT 03-FEB-1999 (first entry)  
 XX DE CD3 delta receptor protein.  
 XX DE Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;  
 KW tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;  
 KW CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;  
 KW protozoan; viral.  
 XX OS Unidentified.  
 XX PA US5843728-A.  
 XX PN 01-DEC-1998.  
 XX PD 05-APR-1995; 95US-0417495.  
 XX PF 06-MAR-1992; 92US-0847566.  
 XX PR 07-MAR-1991; 91US-0665961.  
 XX PR 28-FEB-1994; 94US-0203866.  
 XX PR 05-APR-1995; 95US-0417495.  
 XX PA (GEO) GEN HOSPITAL CORP.  
 XX PI Kolanus W, Romeo C, Seed B;  
 XX WPI; 1999-044582/04.  
 XX DT Membrane-bound chimeric receptors - comprising extracellular portion  
 PT which recognises and binds a target cell and an intracellular  
 PT portion of e.g. a T-cell receptor  
 XX Example 10; Fig 16; 57pp; English.  
 XX CC The present invention describes DNA encoding a membrane-bound chimeric  
 CC receptor comprising: (a) an extracellular portion that specifically  
 CC recognises and binds a target cell or a target infective agent; and (b)  
 CC an intracellular portion of a T-cell receptor CD3, zeta or eta  
 CC polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.  
 CC The present sequence represents a CD3 delta receptor protein from an  
 CC example of the present invention. Cells expressing chimeric receptors of  
 CC the present invention can be administered to mammals in order to destroy  
 CC pathogens (e.g. bacteria, fungi, protozoa or viruses, especially HIV),  
 CC cancer cells or autoimmune-generated cells.  
 XX SQ Sequence 171 AA;  
 Query Match 77.3%; Score 34; DB 20; Length 171;  
 Best Local Similarity 80.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 IIVTDVIATL 10  
 Db 107 iivtdvaitl 116  
 RESULT 7  
 ID R22130 standard; Protein; 275 AA.  
 XX AC R22130;  
 XX DT 22-JUL-1992 (first entry)  
 XX DE

DE Sequence of ovine interleukin IL-2 receptor cDNA.  
 XX Vaccine; antigen; therapeutic agent; immune response enhancer;  
 KW modulator.  
 XX OS Ovis ammon aries.  
 XX PN WO9203574-A.  
 XX PD 05-MAR-1992.  
 XX DT 13-AUG-1991; 91WO-AU000358.  
 XX PF 21-JUN-1991; 91AU-0006840.  
 XX PR 13-AUG-1990; 90AU-0001698.  
 XX PR 11-DEC-1990; 90AU-0003859.  
 XX PR 15-FEB-1991; 91AU-0004621.  
 XX PA (UYME-) UNIV MELBOURNE.  
 XX PA (AWOO) AUSTRALIAN WOOL CORP.  
 XX PI Brandon MR, Andrews AE, Nash AD, Neeusen EN;  
 XX WPI; 1992-096916/12.  
 XX DR N-PSDB; Q22831.  
 XX DT Nucleotide sequences coding for ruminant cytokine(s) or receptors  
 PT - used for producing polypeptide(s) for therapeutic and/or  
 PT adjuvant uses in animals  
 XX Disclosure; Fig 5A; 93pp; English.  
 XX CC The inventors claim a DNA sequence coding for a polypeptide  
 CC exhibiting ruminant cytokine or cytokine receptor activity; ovine  
 CC interleukine (IL)-1-alpha activity; ovine IL-1-alpha activity; ovine  
 CC IL-6 activity; ovine tumour necrosis factor (TNF) alpha activity;  
 CC ovine IL-2 receptor alpha activity; ovine interferon (IFN)-lambda  
 CC activity; or ovine IL-2 activity or homologous sequences, derivs. or  
 CC mutants, or fragments. The recombinant polypeptides are also  
 CC claimed.  
 XX SQ Sequence 275 AA;  
 Query Match 70.5%; Score 31; DB 13; Length 275;  
 Best Local Similarity 85.7%; Pred. No. 1.2e-02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 TDVIATL 10  
 Db 227 tdvvtatl 233  
 RESULT 8  
 ID B41868 standard; Protein; 308 AA.  
 XX AC B41868;  
 XX DT 08-FEB-2001 (first entry)  
 XX DE Human ORFX ORF1632 polypeptide sequence SEQ ID NO:3264.  
 XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.

PN WO200058473-A2.

XX 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

PA Shimkets RA, Leach M;

XX N-PSDB; C76077.

DR Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 2464-2465; 5507pp; English.

CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiparkinsonian; nootropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 308 AA;

Query Match 70.5%; Score 31; DB 21; Length 308;

Best Local Similarity 85.7%; Pred. No. 1.4e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 TDVATL 10

DB 154 tdvwtl 160

RESULT 9

ID Y70723 standard; Protein; 332 AA.

XX Y70723;

XX Y70723;

DT 24-JUL-2000 (first entry)

DE Klebsiella oxytoca yiaK protein.

XX

KW YiaK-S operon; YiaK; YiaJ; regulatory sequence; YiaL; ORF1; YiaX2; LyxK;  
 KW YiaQ; YiaR; YiaS; carbohydrate utilisation; screening; metabolic pathway;  
 KW biological synthesis; anti-infective; antibacterial.

XX Klebsiella oxytoca.

PN WO200022170-A1.

XX 20-APR-2000.

PF 12-OCT-1999; 99WO-US23862.

XX 14-OCT-1998; 98US-0172952.

XX (MICR-) MICROGENOMICS INC.

PI Hoch J, Dartois V;

XX WPI; 2000-318013/27.

DR N-PSDB; Z52319, 52321.

XX Screening for metabolic pathways, useful to provide for the biological  
 PT production of chemicals, antibacterials and other anti-infectives,  
 PT using cells which provide a signal in the presence of a compound  
 PT produced by the pathway -

XX Claim 36; Page 101-103; 137pp; English.

CC The present sequence is the Klebsiella oxytoca yiaK protein, of the  
 CC YiaK-S operon. It includes the genes encoding the Yia operon-related  
 CC polypeptides, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR and YiaS. It is  
 CC preceded by the transcriptional regulatory sequence YiaJ. The function  
 CC of the Yia operon gene products are unknown. It is homologous to YiaK-S  
 CC operon of Escherichia coli and Haemophilus influenzae, which may be  
 CC involved in carbohydrate utilisation. This operon is used in a method for  
 CC screening nucleotide sequences, the products of which can convert a  
 CC source compound to a target compound, using cells which can provide a  
 CC detectable signal in the presence of the target compound. This sequence  
 CC is used to screen and identify biological pathways, that can be used for  
 CC the biological synthesis of chemicals, antibacterials and other  
 CC anti-infectives.

XX Sequence 332 AA;

Query Match 70.5%; Score 31; DB 21; Length 332;

Best Local Similarity 77.8%; Pred. No. 1.5e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 IVTDVIATL 10

DB 231 ivldmiatl 239

RESULT 10

ID B26448

XX B26448 standard; Protein; 397 AA.

AC B26448;

XX 23-FEB-2001 (first entry)

XX Drosophila melanogaster odorant receptor DOR62.

XX Odorant receptor; fruit fly; DOR62; odour recognition; pest control.

XX Drosophila melanogaster.

XX WO200050566-A2.

XX 31-AUG-2000.

XX

PF 25-FEB-2000; 2000WO-US04995.  
 XX  
 PR  
 XX  
 XX  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 XX  
 PI Vosshall LB, Amrein HO, Axel R;  
 XX  
 DR WPI: 2000-572081/53.  
 DR N-PSDB; A94846.  
 XX  
 XX Novel nucleic acid encoding an insect odorant receptor, for identifying  
 PT modulator compounds that are useful in controlling pest population -  
 XX  
 XX Disclosure; Page 76; 176pp; English.  
 XX  
 CC The present sequence is the previously identified Drosophila melanogaster  
 CC odorant receptor DOR62. The odorant genes and proteins, such as those  
 CC provided by the invention, are useful as they aid in the study of the  
 CC olfactory organ in mammals, as well as aiding the understanding of the  
 CC link between odour recognition and behaviour in insects. They also enable  
 CC the identification of compounds capable of activating and inhibiting the  
 CC receptors, allow the control of pest populations via the use of alarm  
 CC odour ligands and via the use of ligands which interfere with the  
 CC interaction between odorant ligands and receptors associated with  
 CC fertility.  
 XX  
 SQ Sequence 397 AA;

Query Match 70.5%; Score 31; DB 21; Length 397;  
 Best Local Similarity 50.0%; Pred. NO. 1.8e+02;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10  
 Db 76 ititdivanl 85

RESULT 11

B20915  
 ID B20915 standard; Protein; 398 AA.

AC B20915;

XX 06-DEC-2000 (first entry)

XX Drosophila odorant receptor DOR 2F.1.

XX Odorant receptor; Drosophila; olfactory receptor;

KW G protein-coupled receptor; GPCR superfamily; transgenic insect;

KW Insect behaviour modification; pest control; pollinator attraction;

KW biosensor; odour detection; odour identification; apiculture.

XX Drosophila melanogaster.

XX WO2000043410-A2.

XX 27-JUL-2000.

XX 25-JAN-2000; 2000WO-US01823.

PR 25-JAN-1999; 99US-0117132.

XX (UYA ) UNIV YALE.

XX Carlson JR, Kim J, Clyne PJ, Warr CG;

XX WPI: 2000-543246/49.

DR N-PSDB; A72250.

XX New nucleic acid encoding a Drosophila olfactory receptor, useful for  
 PT identifying modulating agents -

XX Claim 12; Fig 3A-E; 303pp; English.  
 PS  
 XX Sequences B20901-B20949 represent Drosophila melanogaster odorant  
 CC receptors. These proteins function as olfactory receptors, and  
 CC are thought to be members of the G protein-coupled receptor  
 CC (GPCR) superfamily, which is characterised by the presence of 7  
 CC transmembrane helices. Nucleic acids encoding the Drosophila odorant  
 CC receptors may be used to generate expression constructs, host cells  
 CC containing such constructs, and transgenic insects. Cells which express  
 CC the odorant receptor genes may be used in methods to identify agents  
 CC which modulate expression of these genes, and in methods to identify  
 CC receptor binding partners. The Drosophila odorant receptor nucleic acids  
 CC may also be used to identify corresponding genes in other insects,  
 CC such as those which damage crops or transmit disease. The odorant  
 CC receptor proteins may be used to identify agents which modulate their  
 CC activity, to identify binding partners, as antigens to raise antibodies,  
 CC and in methods to modify insect behaviour. The proteins may be also  
 CC be used in methods of behaviour modification. Such methods may be used to  
 CC study or modify insect behaviour in response to odorants such as  
 CC pheromones. Modification of insect behaviour has a wide range of  
 CC applications, such as in pest control (e.g., by disrupting the feeding or  
 CC mating behaviours of pest species), or for enhancing plant pollination  
 CC (by attracting pollinator species). Odorant receptor proteins and/or  
 CC nucleotides may also be used to identify appetite suppressants, to trap  
 CC odours of a specific type, as biosensors for the detection of explosives,  
 CC drugs, perfumes or pollutants, and in apiculture to modify the behaviour  
 CC of bees, for example, to increase the production of royal jelly.  
 XX  
 SQ Sequence 398 AA;

Query Match 70.5%; Score 31; DB 21; Length 398;

Best Local Similarity 50.0%; Pred. NO. 1.8e+02;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10  
 Db 76 ititdivanl 85

RESULT 12

G30315

ID G30315 standard; Protein; 421 AA.

XX G30315;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 36219.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126284.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.



PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	30-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	07-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	11-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145192.
PR	14-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134221.	PR	26-JUL-1999;	99US-0145276.
PR	18-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145913.
PR	19-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145918.
PR	20-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145919.
PR	21-MAY-1999;	99US-0135124.	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146386.
PR	25-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146388.
PR	27-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-0146389.
PR	28-MAY-1999;	99US-0136782.	PR	03-AUG-1999;	99US-0147038.
PR	01-JUN-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147204.
PR	03-JUN-1999;	99US-0137528.	PR	04-AUG-1999;	99US-0147302.
PR	04-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147192.
PR	07-JUN-1999;	99US-0137724.	PR	05-AUG-1999;	99US-0147260.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147303.
PR	10-JUN-1999;	99US-0138540.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147493.
PR	14-JUN-1999;	99US-0139119.	PR	09-AUG-1999;	99US-0147935.
PR	16-JUN-1999;	99US-0139452.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139453.	PR	11-AUG-1999;	99US-0148319.
PR	17-JUN-1999;	99US-0139492.	PR	12-AUG-1999;	99US-0148341.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139455.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139456.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139457.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139458.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139461.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139463.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139750.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139763.	PR	26-AUG-1999;	99US-0150884.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151065.
PR	22-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151066.
PR	23-JUN-1999;	99US-0140353.	PR	27-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0140354.	PR	30-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	99US-0140695.	PR	31-AUG-1999;	99US-0151438.
PR	28-JUN-1999;	99US-0140823.	PR	01-SEP-1999;	99US-0151930.
PR	29-JUN-1999;	99US-0140991.	PR	07-SEP-1999;	99US-0152363.
PR	30-JUN-1999;	99US-0141287.	PR	10-SEP-1999;	99US-0153070.
PR	01-JUL-1999;	99US-0141287.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141284.	PR	15-SEP-1999;	99US-0154018.
PR	02-JUL-1999;	99US-0142055.	PR	16-SEP-1999;	99US-0154039.
PR	06-JUL-1999;	99US-0142390.	PR	20-SEP-1999;	99US-0154779.
PR	08-JUL-1999;	99US-0142803.	PR	22-SEP-1999;	99US-0155139.
PR	09-JUL-1999;	99US-0142920.	PR	23-SEP-1999;	99US-0155486.
PR	12-JUL-1999;	99US-0142977.	PR	24-SEP-1999;	99US-0155659.
PR	13-JUL-1999;	99US-0143342.	PR	28-SEP-1999;	99US-0156458.
PR	14-JUL-1999;	99US-0143624.	PR	29-SEP-1999;	99US-0156596.
PR	15-JUL-1999;	99US-0144005.	PR	04-OCT-1999;	99US-0157117.
PR	16-JUL-1999;	99US-0144086.	PR	05-OCT-1999;	99US-0157753.
PR	16-JUL-1999;	99US-0144086.	PR	06-OCT-1999;	99US-0157865.
PR	19-JUL-1999;	99US-0144325.	PR	07-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144331.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144332.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.

PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.5%; Score 31; DB 21; Length 421;

Best Local Similarity 50.0%; Pred. No. 2e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10  
 :::||: |||  
 Db 31 mvtldlaatl 40

RESULT 13  
 G30314  
 ID G30314 standard; Protein; 428 AA.

AC G30314;  
 XX XX  
 DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 36218.

XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 29-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 25-MAY-1999; 99US-0136392.  
 PR 27-MAY-1999; 99US-0136782.  
 PR 28-MAY-1999; 99US-0137222.  
 PR 01-JUN-1999; 99US-0137528.  
 PR 03-JUN-1999; 99US-0137502.  
 PR 04-JUN-1999; 99US-0137724.  
 PR 07-JUN-1999; 99US-0138094.  
 PR 08-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.

```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145091.
PR 22-JUL-1999; 99US-0145092.
PR 22-JUL-1999; 99US-0145145.
PR 22-JUL-1999; 99US-0145218.
PR 22-JUL-1999; 99US-0145224.
PR 22-JUL-1999; 99US-0145226.
PR 22-JUL-1999; 99US-0145276.
PR 22-JUL-1999; 99US-0145313.
PR 22-JUL-1999; 99US-0145318.
PR 22-JUL-1999; 99US-0145319.
PR 22-JUL-1999; 99US-0145351.
PR 22-JUL-1999; 99US-0145356.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147320.
PR 05-AUG-1999; 99US-0147322.
PR 06-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 70.5%; Score 31; DB 21; Length 428;
Best Local Similarity 50.0%; Pred. NO. 2e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10
   :::||: |||
Db 38 mvitdlaatl 47

RESULT 14
G30313
ID G30313 standard; Protein; 475 AA.
AC G30313;
XX
XX 17-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 36217.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
```

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142377.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 13-SEP-1999; 99US-0153070.  
PR 15-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 20-SEP-1999; 99US-0154039.  
PR 22-SEP-1999; 99US-0154779.  
PR 23-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

C97901 to C98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in B35234 to B34006. The human colon cancer antigens can have cytosolic, cardioactive, muscular; neuroprotective, immunomodulatory, gynaecological, gastrointestinal, vulvular, nephrotropic, antineoplastic and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The polynucleotides may be used in diagnostics and research, such as for chromosome identification, and as hybridisation probes. The proteins may also be used to prevent diseases such as neural disorders, immune system disorders, muscular disorders, reproductive disorders.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:15:02 ; Search time 53.33 seconds  
(without alignments)  
24.176 Million cell updates/sec

Title: US-09-202-305-20

Perfect score: 58  
Sequence: 1 SSDVPCDATLT 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_15.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mnc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No: is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	42	72.4	315	10 Q9SI22	Q9SI22 arabidopsis
2	42	72.4	426	11 Q9RD9	Q9RD9 rattus norv
3	38.5	66.4	388	10 Q9LZ00	Q9LZ00 arabidopsis
4	38	65.5	1530	4 Q75763	Q75763 homo sapien
5	37	63.8	151	14 Q11327	Q11327 molluscum c
6	37	63.8	373	10 Q93754	Q93754 arabidopsis
7	37	63.8	406	14 Q98216	Q98216 molluscum c
8	37	63.8	426	11 Q9Z0L0	Q9Z0L0 mus musculu
9	37	63.8	508	14 Q74118	Q74118 human immun
10	37	63.8	713	2 Q9RWI3	Q9RWI3 deinococcus
11	37	63.8	1335	4 Q9P222	Q9P222 homo sapien
12	37	63.8	1813	4 Q9UDR7	Q9UDR7 homo sapien
13	37	63.8	1967	14 Q85652	Q85652 blueberry s
14	37	63.8	4019	4 Q9NKR13	Q9NKR13 homo sapien
15	36	62.1	139	2 Q9RTA8	Q9RTA8 deinococcus
16	36	62.1	302	5 Q9NEG3	Q9NEG3 drosophila
17	36	62.1	327	5 Q9W4V4	Q9W4V4 drosophila
18	36	62.1	384	10 Q42631	Q42631 brassica na
19	36	62.1	527	5 Q9N510	Q9N510 caenorhabdi

20	62.1	636	14	Q90142	Q90142 spodoptera
21	62.1	788	4	Q9UGA8	Q9UGA8 homo sapien
22	62.1	972	10	Q82072	Q82072 triticum ae
23	62.1	1063	14	Q9J844	Q9J844 spodoptera
24	60.3	153	2	Q86586	Q86586 streptomyce
25	60.3	225	4	Q15358	Q15358 homo sapien
26	60.3	473	10	Q22562	Q22562 arabidopsis
27	60.3	527	10	Q9LRC8	Q9LRC8 scutellaria
28	60.3	607	10	Q96536	Q96536 arabidopsis
29	60.3	633	11	P70419	P70419 mus musculu
30	60.3	783	14	Q9WJ22	Q9WJ22 ophiostoma
31	60.3	1007	5	Q22286	Q22286 caenorhabdi
32	60.3	1029	5	Q9N9C2	Q9N9C2 leishmania
33	60.3	1060	5	Q9VXG7	Q9VXG7 drosophila
34	60.3	1374	5	Q9VXG6	Q9VXG6 drosophila
35	60.3	1430	5	Q9VAV3	Q9VAV3 drosophila
36	60.3	2230	5	Q9VAV4	Q9VAV4 drosophila
37	58.6	112	5	O45700	O45700 caenorhabdi
38	58.6	188	2	P73952	P73952 synechocyst
39	58.6	327	2	O53285	O53285 mycobacteri
40	58.6	365	5	Q23018	Q23018 caenorhabdi
41	58.6	394	2	O52577	O52577 pseudomonas
42	58.6	469	4	Q9NSS3	Q9NSS3 homo sapien
43	58.6	469	4	Q9NSS2	Q9NSS2 homo sapien
44	58.6	613	13	Q03711	Q03711 xenopus lae
45	58.6	614	5	Q9N924	Q9N924 trypanosoma

#### ALIGNMENTS

```

RESULT 1
Q9SI22
AC Q9SI22 PRELIMINARY; PRT: 315 AA.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL 35.5 KDA PROTEIN.
GN AT2G05060.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-L., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umavay L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL; AC007211; AAD25594.1; -
DR HSP; P08631; IAD5
DR INTERPRO; IPR000719; -
DR INTERPRO; IPR001245; -
DR INTERPRO; IPR002290; -
DR PFAM; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 315 AA; 35499 MW; 813DBIC2EDF74AEB CRC64;

```

Query Match 72.4%; Score 42; DB 10; Length 315;  
Best Local Similarity 70.0%; Pred. No. 2.1;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSDVPCDATL 10  
 ||:|||||  
 Db 241 SKDIPCDATL 250

RESULT 2  
 Q9QYD9 PRELIMINARY; PRT; 426 AA.

AC Q9QYD9;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE 5T4 ONCORETAL ANTIGEN HOMOLOG.  
 GN 5T4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CEREBELLUM;  
 RA Ninkina N.N., Buchman V.L.;  
 RT "Structure and expression of the rat 5T4 gene."  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF063939; AAF21770.1;  
 DR INTERPRO: IPR000372;  
 DR INTERPRO: IPR000483;  
 DR INTERPRO: IPR001611;  
 DR PFAM: PF00560; LRR; 6.  
 DR PFAM: PF01462; LRRNT; 1.  
 DR PFAM: PF01463; LRRCT; 1.  
 DR PRINTS: PR00019; LEURICHRPT.  
 SQ SEQUENCE 426 AA; 46439 MW; 4EEFF7DA86B545B0 CRC64;

Query Match 72.4%; Score 42; DB 11; Length 426;  
 Best Local Similarity 80.0%; Pred. No. 2.8;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSDVPCDATL 10  
 ||:|||||  
 Db 345 SSDLDCATL 354

RESULT 3  
 Q9LZU0 PRELIMINARY; PRT; 388 AA.

AC Q9LZU0;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE HYPOTHETICAL 45.2 KDA PROTEIN.  
 GN F16L2.110.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unseid M., Mewes H.W.,  
 RA Rued S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL162459; CAB82814.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 388 AA; 45186 MW; F92E00CA6D3AD68A CRC64;

Query Match 66.4%; Score 38.5; DB 10; Length 388;  
 Best Local Similarity 90.0%; Pred. No. 12;  
 Matches 9; Conservative 0; Mismatches 1; Gaps 1;

QY 1 SSDVPCDATL 10  
 ||:|||||  
 Db 274 SSD-PCDATL 282

RESULT 4  
 Q75763 PRELIMINARY; PRT; 1530 AA.

AC Q75763;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE SYNAPTONEMAL COMPLEX LATERAL ELEMENT PROTEIN.  
 GN SCP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98256424; PubMed=9592139;  
 RA Offenberger H.H., Schalk J.A.C., Meuwissen R.L.J., van Aalderen M.,  
 RA Kester H.A., Dietrich A.J.J., Heyting C.;  
 RT "SCP2: a major protein component of the axial elements of synaptonemal  
 complexes of the rat."  
 RL Nucleic Acids Res. 26:2572-2579(1998).  
 DR EMBL: Y08982; CAA70171.1;  
 SQ SEQUENCE 1530 AA; 175710 MW; 6B6D7363CA3171F8 CRC64;

Query Match 65.5%; Score 38; DB 4; Length 1530;  
 Best Local Similarity 85.7%; Pred. No. 57;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DVPCDAT 9  
 ||:|||||  
 Db 1268 DMPCDAT 1274

RESULT 5  
 O11327 PRELIMINARY; PRT; 151 AA.

AC O11327;  
 DT 01-JUL-1997 (Tremblrel. 04, Created)  
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
 DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)  
 DE CLONE HI-31 HOMOLOG OF VACCINIA I6L (HI-31) (FRAGMENT).  
 GN HI-31.  
 OS Molluscum contagiosum virus subtype 1 (MCV1).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Molluscipoxvirus.  
 OX NCBI\_TaxID=10280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Moratilla M., Agromayor M., Nunez A., Funes J.M., Varas A.J.,  
 RA Lopez-Esteban J.L., Esteban M., Martin-Gallardo A.;  
 RL Virus Genes 0:0-0(0).  
 DR EMBL: U86907; AAB57959.1;  
 FT NON-TER 1 151  
 FT NON-TER 151 151  
 SQ SEQUENCE 151 AA; 16303 MW; E1175252F696858E CRC64;

Query Match 63.8%; Score 37; DB 14; Length 151;  
 Best Local Similarity 66.7%; Pred. No. 9.3;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 DVPCDATLT 11



Db 43 DLPCDTELT 51  
|:|:| |

## RESULT 6

ID P93754 PRELIMINARY; PRT; 373 AA.  
AC P93754;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE SER/THR KINASE ISOLOG.  
GN T06D20.19.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 11;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Rounsley S.D., Lin X., Ketchum K.A., Phillips C.A., Brandon R.C.,  
RA Fuhrmann J.L., White O., Kerlavage A.R., Adams M.D., Somerville C.R.,  
RA Venter J.C.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U90439; AAB63551.1; -.  
DR MENDEL; 14302; Arath;2338;14302.  
DR INTERPRO; IPR000719; -.  
DR INTERPRO; IPR002290; -.  
DR PFAM; PF00069; Pkinase; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
SQ SEQUENCE 373 AA; 42910 MW; 4F41D036BFBF8A7F8 CRC64;

## Query Match

Best Local Similarity 63.8%; Score 37; DB 10; Length 373;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## QY 2 SDVPCDA 8

Db 243 SLDPCDA 249  
|:|:| |

## RESULT 7

ID Q98216 PRELIMINARY; PRT; 406 AA.  
AC Q98216;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE MC048L.  
GN MC048L.  
OS Molluscum contagiosum virus subtype 1 (MCV1).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Molluscipoxvirus.  
OX NCBI\_TaxID=10280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96325459; PubMed=8670425;  
RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,  
RA Moss B.;  
RT "Genome sequence of a human tumorigenic poxvirus: prediction of  
specific host response-evasion genes.";  
RL Science 273:813-816(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,  
RA Moss B.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U60315; AAC55176.1; -.  
SQ SEQUENCE 406 AA; 44568 MW; 6AC286B5DBE8CD42 CRC64;

Query Match 63.8%; Score 37; DB 14; Length 406;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

## QY 3 DVPCDATLT 11

Db 286 DLPCDTELT 294  
|:|:| |

## RESULT 8

Q9Z0L0 PRELIMINARY; PRT; 426 AA.  
ID Q9Z0L0;  
AC Q9Z0L0;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE 5T4 ONCOFETAL TROPHOBLAST GLYCOPROTEIN PRECURSOR.  
GN TPBG OR 5T4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RA King K.W., Sheppard F.C., Westwater C., Stern P.L., Myers K.A.;  
RT "Genomic organisation of the mouse and human 5T4 oncofetal leucine-  
rich glycoprotein gene and expression in foetal and adult murine  
tissues.";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ012160; CAA09931.1; -.  
DR MGD; MGI:1341264; Tpbq.  
DR INTERPRO; IPR000372; -.  
DR INTERPRO; IPR000483; -.  
DR INTERPRO; IPR001611; -.  
DR PFAM; PF00560; LRR; 6.  
DR PFAM; PF01462; LRRNT; 1.  
DR PFAM; PF01463; LRRCT; 1.  
DR PRINTS; PR00019; LEURICHRPT.  
KW Signal.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 426 5T4 ONCOFETAL TROPHOBLAST GLYCOPROTEIN.  
SQ SEQUENCE 426 AA; 46514 MW; C43B966A07ED9EF5 CRC64;

Query Match 63.8%; Score 37; DB 11; Length 426;  
Best Local Similarity 70.0%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

## QY 1 SSDVPCDATL 10

Db 345 SSDLDCDAVL 354  
|:|:| |

## RESULT 9

Q74118 PRELIMINARY; PRT; 508 AA.  
ID Q74118;  
AC Q74118;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE GP105 (FRAGMENT).  
GN ENV.  
OS Human immunodeficiency virus type 2.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11709;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HIV2D868;  
RX MEDLINE=94187055; PubMed=8139000;  
RA Grez M., Dietrich U., Balfe P., Von Briesen H., Maniar J.K.,  
RA Mahambre G., Delwart E.L., Mullins J.I., Ruebsamen-Waigmann H.;

RT \*Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1  
 RT and HIV-2) mixed infections in India reveals a recent spread of HIV-1  
 RT and HIV-2 from a single ancestor for each of these viruses.";  
 RL J. Virol. 68:2161-2168(1994).

DR EMBL; U07108; AAA17669.1; -.  
 DR INTERPRO; IPR000777; -.  
 DR PFAM; PF00516; GP120; 1.  
 FT NON\_TER 508 508  
 SQ SEQUENCE 508 AA; 58165 MW; 9B103784A08DCB9C CRC64;

Query Match 63.8%; Score 37; DB 14; Length 508;  
 Best Local Similarity 72.7%; Pred. No. 30;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSDVPCDATT 11  
 Db 178 SSDVCEATGT 188  
 ||||| |||||

RESULT 10  
 Q9RWI3 PRELIMINARY; PRT; 713 AA.  
 AC Q9RWI3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE HYPOTHETICAL 73.9 KDA PROTEIN.  
 GN DR0685.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RL;  
 RA MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome Sequence of the Radioresistant Bacterium Deinococcus  
 RT radiodurans RL.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL; AE001925; AAF10267.1; -.  
 DR TIGR; DR0685; -.  
 DR INTERPRO; IPR002048; -.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 713 AA; 73881 MW; 8FF76CC6D79B5B4E CRC64;

Query Match 63.8%; Score 37; DB 2; Length 713;  
 Best Local Similarity 77.8%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 DVPCDATT 11  
 Db 224 DVPCATST 232  
 ||||| |||||

RESULT 11  
 Q9P222 PRELIMINARY; PRT; 1335 AA.  
 AC Q9P222;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE KIAA1506 PROTEIN (FRAGMENT).  
 GN KIAA1506.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human  
 RT genes XVII. The complete sequences of 100 new cDNA clones from brain  
 RT which code for large proteins in vitro.";  
 RL DNA Res. 7:143-150(2000).  
 DR EMBL; AB040939; BAA96030.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 1335 1335  
 SQ SEQUENCE 1335 AA; 146816 MW; 42AA56A161FA2115 CRC64;

Query Match 63.8%; Score 37; DB 4; Length 1335;  
 Best Local Similarity 85.7%; Pred. No. 78;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSDVPCD 7  
 Db 1188 SSDLPCD 1194  
 ||||| |||||

RESULT 12  
 Q9UDR7 PRELIMINARY; PRT; 1813 AA.  
 ID Q9UDR7;  
 AC Q9UDR7;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE WUGSC.H.DJ0981007.3 PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99063792; PubMed=9847074;  
 RA Sulston J.E., Waterston R.;  
 RT "Toward a complete human genome sequence.";  
 RL Genome Res. 8:1097-1108(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Joshi C., McGahey B.;  
 RT "The sequence of Homo sapiens PAC clone RP5-98107.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.H.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006017; AAD45822.1; -.  
 DR INTERPRO; IPR001214; -.  
 DR PFAM; PF00856; SET; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 1813 AA; 201257 MW; F13A1EDCA321F40A CRC64;

Query Match 63.8%; Score 37; DB 4; Length 1813;  
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSDVPCD 7  
 Db 304 SSDLPCD 310  
 ||||| |||||

RESULT 13

Q65652  
 ID Q65652 PRELIMINARY; PRT; 1967 AA.  
 AC Q65652;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE BLUEBERRY SCORCH CARLAVIRUS MRNA.  
 OS BLUEBERRY SCORCH VIRUS.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=31722;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NJ-2;  
 RX MEDLINE=94201746; PubMed=8151289;  
 RA Cavilleer T.D., Halpern B.T., Lawrence D.M., Podleckis E.V.,  
 RA Martin R.R., Hillman B.I.;  
 RT "Nucleotide sequence of the carlavirus associated with blueberry  
 RT scorch and similar diseases.";  
 RL J. Gen. Virol. 75:711-720(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NJ-2;  
 RX MEDLINE=95176530; PubMed=7871721;  
 RA Lawrence D.M., Rozanov M.N., Hillman B.I.;  
 RT "Autocatalytic processing of the 223-kDa protein of blueberry scorch  
 RT carlavirus by a papain-like proteinase.";  
 RL Virology 207:127-135(1995).  
 DR EMBL; L25658; AAA68984.1; -.  
 DR MEGP; C23.001; -.  
 SQ SEQUENCE 1967 AA; 223428 MW; 4CEDD6AF14980188 CRC64;

Query Match 63.8%; Score 37; DB 14; Length 1967;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSDVPCDA 8  
 :||| ||||  
 DB 661 TSDAPCDA 668

RESULT 14  
 Q9NR13  
 ID Q9NR13 PRELIMINARY; PRT; 4019 AA.  
 AC Q9NR13;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE ALR-LIKE PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CERVICAL CARCINOMA;  
 RA Chow V.T.K., Tan Y.C.;  
 RT "ALR-like protein, a novel human cDNA whose product is homologous to  
 RT the ALR protein.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF264750; AAF74766.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 4019 AA; 442853 MW; 772B1A7223C4E716 CRC64;

Query Match 63.8%; Score 37; DB 4; Length 4019;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSDVPCD 7  
 :|||: ||||  
 DB 2510 SSDLPCD 2516

RESULT 15  
 Q9RTA8  
 ID Q9RTA8 PRELIMINARY; PRT; 139 AA.  
 AC Q9RTA8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE ORGANIC HYDROPEROXIDE RESISTANCE PROTEIN.  
 DR1857.  
 GN Deinococcus radiodurans.  
 OS Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome Sequence of the Radioresistant Bacterium Deinococcus  
 RT radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL; AE002025; AAF11408.1; -.  
 DR TIGR; DR1857; -.  
 SQ SEQUENCE 139 AA; 14530 MW; DC95B02CF492642E CRC64;

Query Match 62.1%; Score 36; DB 2; Length 139;  
 Best Local Similarity 66.7%; Pred. No. 13;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 DVPDCAFLT 11  
 :||| :|||  
 DB 72 DVPADSTIT 80

Search completed: May 10, 2001, 10:15:04  
 Job time: 263 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:15:24 ; Search time 20.13 seconds  
(without alignments)  
18.719 Million cell updates/sec

Title: US-09-202-305-20  
Perfect score: 58  
Sequence: 1 SSDVPCDATLT 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	100.0	138	1	TCA_MOUSE
2	37	63.8	329	1	IHA_CHICK
3	36	62.1	290	1	DPSD_HAEIN
4	36	62.1	960	1	CAP2_SORBI
5	36	62.1	967	1	CAP2_MAIZE
6	35	60.3	960	1	CAP2_MESCR
7	34	58.6	136	1	TVC_HUMAN
8	34	58.6	257	1	ETFB_BACSU
9	34	58.6	371	1	Y816_SYNY3
10	34	58.6	387	1	ARRC_XENLA
11	34	58.6	964	1	CAPP_AMAHP
12	34	58.6	1780	1	POLG_MVEV
13	34	58.6	3430	1	POLG_WNV
14	34	58.6	3432	1	POLG_JAEV1
15	34	58.6	3432	1	POLG_JAEV5
16	34	58.6	3432	1	POLG_JAEVJ
17	34	58.6	3433	1	POLG_KUNJM
18	33	56.9	364	1	YS80_MYCTU
19	33	56.9	467	1	MM08_HUMAN
20	33	56.9	511	1	GUNB_PSEFL
21	33	56.9	583	1	T2F1_FIAOK
22	33	56.9	588	1	GRN_RAT
23	33	56.9	690	1	AFT1_YEAST
24	33	56.9	899	1	Y188_HUMAN
25	33	56.9	1549	1	Y1R3_YEAST
26	33	56.9	1549	1	YJW2_YEAST
27	33	56.9	3358	1	PGCV_MOUSE
28	33	56.9	3358	1	PGCV_HUMAN
29	33	56.9	3562	1	PGCV_CHICK
30	32	55.2	171	1	YSYM_RHET
31	32	55.2	211	1	END3_TREPA
32	32	55.2	228	1	DHA9_POLMI
33	32	55.2	240	1	RR2_EUGER

34	32	55.2	276	1	YRD4_CAEEL	Q09574 caenorhabdi
35	32	55.2	376	1	P53_ICTPU	O93379 ictalurus p
36	32	55.2	421	1	PGUR_MEDSA	O40312 medicago sa
37	32	55.2	436	1	TBX6_HUMAN	O93947 homo sapien
38	32	55.2	457	1	CD4_RAT	P05540 rattus norv
39	32	55.2	461	1	YAU2_SCHPO	Q10159 schizosacch
40	32	55.2	540	1	TBX6_MOUSE	P70327 mus musculu
41	32	55.2	589	1	GRN_MOUSE	P28798 mus musculu
42	32	55.2	593	1	GRN_HUMAN	P28799 h granullins
43	32	55.2	633	1	P2AB_BACTK	P21254 bacillus th
44	32	55.2	633	1	C2AD_BACTU	Q9rmg3 bacillus th
45	32	55.2	634	1	HWPI_CANAL	P46593 candida alb

ALIGNMENTS

RESULT 1	TCA_MOUSE	STANDARD;	PRT; 138 AA.
ID	P01849;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	21-JUL-1986 (Rel. 01, Last annotation update)		
DE	T-CELL RECEPTOR ALPHA CHAIN C REGION.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A. (CLONE PY14).		
RC	STRAIN=BALB.B;		
RX	MEDLINE=85036635; PubMed=6208487;		
RA	Saito H., Kranz D.M., Takagaki Y., Hayday A.C., Eisen H.N.,		
RA	Tonegawa S.;		
RT	"A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes.";		
RL	Nature 312:36-40(1984).		
RN	[2]		
RP	SEQUENCE FROM N.A. (HYBRIDOMA 2B4).		
RX	MEDLINE=85036634; PubMed=6548551;		
RA	Chien Y., Becker D.M., Lindsten T., Okamura M., Cohen D.I.,		
RA	Davis M.M.;		
RT	"A third type of murine T-cell receptor gene.";		
RL	Nature 312:31-35(1984).		
CC	-!- MISCELLANEOUS: CLONE PHDS58 WAS ISOLATED FROM A CYTOTOXIC T LYMPHOCYTE.		
DR	PIR: A02132; RWMSC8.		
KW	T-cell; Receptor; Transmembrane; Glycoprotein.		
FT	NON_TER 1 1		
FT	DOMAIN 1 112		
FT	TRANSMEM 113 133		
FT	DOMAIN 134 138		
FT	CARBOHYD 68 68		
FT	CARBOHYD 82 82		
FT	CARBOHYD 111 111		
FT	VARIANT 2 2		
SQ	SEQUENCE 138 AA: 15494 MW; 984B77953AA80444 CRC64;		

Query Match 100.0%; Score 58; DB 1; Length 138;  
Best Local Similarity 100.0%; Pred. No. 0.00045;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	SSDVPCDATLT 11
Db	87	SSDVPCDATLT 97

RESULT 2  
IHA\_CHICK STANDARD; PRT; 329 AA.  
ID IHA\_CHICK  
AC P43031; Q90708;

DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	PHOSPHATIDYL-SERINE DECARBOXYLASE PROENZYME (EC 4.1.1.65).
OS	PSD OR H10160.
OC	Haemophilus influenzae.
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC	Haemophilus.
OX	NCBI_TaxID=727;
RN	[1]
RP	SEQUENCE FROM N.A.
RRP	STRAIN-RD / KW20 / ATCC 51907;
RC	Barcak G.J., Helmer S.R.;
RA	Submitted (JAN-1995), to the EMBL/GenBank/DDBJ databases.
RRL	[2]
RP	SEQUENCE FROM N.A.
RRP	STRAIN-RD / KW20 / ATCC 51907;
RRX	MEDLINE=95350630; PubMed=7542800;
RR	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA	Fire L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.H.,
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA	Venter J.C.;
RT	"Whole-genome random sequencing and assembly of Haemophilus
RT	influenzae RD.";
RL	Science 269:496-512(1995).
CC	-!- CATALYTIC ACTIVITY: PHOSPHATIDYL-L-SERINE =
CC	PHOSPHATIDYLETHANOLAMINE + CO(2).
CC	-!- COFACTOR: REQUIRES A PYRUVYL GROUP FOR ITS ACTIVITY
CC	(BY SIMILARITY).
CC	-!- SUBUNIT: MULTIMER OF THE HETERODIMER (BY SIMILARITY).
CC	-!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (BY SIMILARITY).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to licens@isb-sib.ch).
CC	-----
QR	EMBL; U20229; AAC62138.1; -;
DR	EMBL; U32701; AAC21829.1; -;
DR	TIGR; H10160; -;
KW	Phospholipid biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen.
FT	CHAIN 1 256
FT	CHAIN 257 290
FT	CHAIN (BY SIMILARITY).
FT	PHOSPHATIDYL-SERINE DECARBOXYLASE ALPHA
FT	CHAIN (BY SIMILARITY).
FT	CLEAVAGE (NONHYDROLYTICAL)
FT	(BY SIMILARITY).
FT	MOD_RES 257 257
FT	CONVERTED TO A PYRUVYL GROUP
FT	(BY SIMILARITY).
SQ	SEQUENCE 290 AA; 33197 MW; 66766BC16AC82B2D CRC64;
QY	Query Match 62.1%; Score 36; DB 1; Length 290;
Best Local Similarity 85.7%;	
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	
DB	154 VPCDATL 160
QY	4 VPCDATL 10
:	
ID	CAP2_SORBI STANDARD; PRT; 960 AA.
AC	P29194;
DT	01-DEC-1992 (Rel. 24, Created)
DT	01-DEC-1992 (Rel. 24, Last sequence update)
DT	01-DEC-1992 (Rel. 24, Last sequence update)

30-MAY-2000 (Rel. 39, Last annotation update)  
 PHOSPHOENOLPYRUVATE CARBOXYLASE 2 (EC 4.1.1.31) (PEPCASE) (CP28).  
 Sorghum bicolor (Sorghum) (Sorghum vulgare).  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;  
 Andropogoneae; Sorghum.  
 NCBI\_TaxID=4558;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=92032766; PubMed=1840686;  
 Lepiniec L., Santi S., Keryer E., Amiet V., Vidal J., Gadal P.,  
 Cretin C.;  
 "Complete nucleotide sequence of one member of the Sorghum  
 phosphoenolpyruvate carboxylase gene family.";  
 Plant Mol. Biol. 17:1077-1079(1991).  
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID  
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.  
 CC -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALOACETATE = H(2)O +  
 CC PHOSPHOENOLPYRUVATE + CO(2).  
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN  
 CC C3 PHOTOSYNTHESIS.  
 CC -!- SUBUNIT: HOMOTETRAMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X59925; CAA42549.1;  
 CC PIR; S18240; S18240.  
 CC InterPro: IPR001449;  
 CC Pfam; PF00311; PEPCase; 1.  
 CC PRINTS; PR00150; PEPCARBXLASE.  
 CC PROSITE; PS00393; PEPCASE\_2; 1.  
 CC PROSITE; PS00781; PEPCASE\_1; 1.  
 CC KW Lyase; Carbon dioxide fixation; Allosteric enzyme; Multigene family;  
 CC Tricarboxylic acid cycle; Photosynthesis.  
 CC ACT\_SITE 167 167 BY SIMILARITY.  
 CC ACT\_SITE 595 595 BY SIMILARITY.  
 CC SEQUENCE 960 AA; 109547 MW; 65FDC9B71156B18 CRC64;

Query Match 62.1%; Score 36; DB 1; Length 960;  
 Best Local Similarity 70.0%; Pred. No. 40;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SDVPCDATT 11  
 ||| :|||  
 Db 387 SDIPEATT 396

RESULT 5  
 CAP2\_MAIZE STANDARD; PRT; 967 AA.  
 AC P51059;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE 2 (EC 4.1.1.31) (PEPCASE).  
 GN PEPA OR PEP.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; PACC clade; Panicoideae;  
 OC Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. H84; TISSUE=Root;

RX MEDLINE=93054411; PubMed=1429504;  
 RA Kawanura T., Shigesada K., Toh H., Okumura S., Yanagisawa S.,  
 Izui K.;  
 "Molecular evolution of phosphoenolpyruvate carboxylase for C4  
 photosynthesis in maize: comparison of its cDNA sequence with a newly  
 RT isolated cDNA encoding an isozyme involved in the anaplerotic  
 RT function.";  
 RL J. Biochem. 112:147-154(1992).  
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID  
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.  
 CC -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALOACETATE = H(2)O +  
 CC PHOSPHOENOLPYRUVATE + CO(2).  
 CC -!- ENZYME REGULATION: BY LIGHT-REVERSIBLE PHOSPHORYLATION (BY  
 CC SIMILARITY).  
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN  
 CC C3 PHOTOSYNTHESIS.  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X61489; CAA43709.1;  
 CC MaizeDB; 30066;  
 CC InterPro: IPR001449;  
 CC Pfam; PF00311; PEPCase; 1.  
 CC PRINTS; PR00150; PEPCARBXLASE.  
 CC PROSITE; PS00393; PEPCASE\_2; 1.  
 CC PROSITE; PS00781; PEPCASE\_1; 1.  
 CC KW Lyase; Carbon dioxide fixation; Allosteric enzyme; Multigene family;  
 CC Tricarboxylic acid cycle; Photosynthesis.  
 CC MOD\_RES 13 13 PHOSPHORYLATION (BY SIMILARITY).  
 CC ACT\_SITE 174 174 BY SIMILARITY.  
 CC ACT\_SITE 602 602 BY SIMILARITY.  
 CC SEQUENCE 967 AA; 109998 MW; 7034A2AD5521645B CRC64;

Query Match 62.1%; Score 36; DB 1; Length 967;  
 Best Local Similarity 70.0%; Pred. No. 40;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SDVPCDATT 11  
 ||| :|||  
 Db 394 SDIPEATT 403

RESULT 6  
 CAP2\_MESCR STANDARD; PRT; 960 AA.  
 AC P16097;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE 2 (EC 4.1.1.31) (PEPCASE).  
 GN PPC2.  
 OS Mesembryanthemum crystallinum (Common ice plant).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;  
 OC Caryophyllales; Aizoaceae; Mesembryanthemum.  
 OX NCBI\_TaxID=3544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89386016; PubMed=2780305;  
 RA Cushman J.C., Bohnert H.J.;  
 RT "Nucleotide sequence of the Ppc2 gene encoding a housekeeping isoform  
 RT of phosphoenolpyruvate carboxylase from Mesembryanthemum  
 crystallinum.";

RL Nucleic Acids Res. 17:6743-6744(1989).  
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID  
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.  
 CC -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALOACETATE = H(2)O +  
 CC PHOSPHENOLPYRUVATE + CO(2).  
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN  
 CC C3 PHOTOSYNTHESIS.  
 CC -!- SUBUNIT: HOMOTETRAMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X14588; CAA32728.2; -.  
 CC PIR; S05550; QYIX2.  
 CC InterPro; IPR001449; -.  
 CC Pfam; PF00311; PEPCase; 1.  
 CC PRINTS; PR00150; PEPCARBLASE.  
 CC PROSITE; PS00393; PEPCASE\_2; 1.  
 CC PROSITE; PS00781; PEPCASE\_1; 1.  
 CC Lyase; Carbon dioxide fixation; Allosteric enzyme; Multigene family;  
 CC Tricarboxylic acid cycle; Photosynthesis.  
 CC ACT\_SITE 165 165 BY SIMILARITY.  
 CC ACT\_SITE 593 593 BY SIMILARITY.  
 CC SEQUENCE 960 AA; 109183 MW; 5318B18E37B76B5A CRC64;  
 CC  
 CC Query Match 60.3%; Score 35; DB 1; Length 960;  
 CC Best Local Similarity 70.08; Pred. No. 62;  
 CC Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 CC  
 CC QY 2 SDVPCDAPLT 11  
 CC ||:| ||| |  
 CC Db 385 SDIPEDATYT 394  
 CC  
 CC RESULT 7  
 CC TVC\_HUMAN STANDARD; PRT; 136 AA.  
 CC AC P03979;  
 CC DT 23-OCT-1986 (Rel. 02, Created)  
 CC DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE T-CELL RECEPTOR GAMMA CHAIN V REGION PT-GAMMA-1/2 PRECURSOR.  
 CC GN TCRB.  
 CC OS Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=86206003; PubMed=3458221;  
 CC RA Djalynas D.P., Murte C., Quettermous T., Boss J.M., Leiden J.M.,  
 CC Seidman J.G., Strominger J.L.;  
 CC "Cloning and sequence analysis of complementary DNA encoding an  
 CC aberrantly rearranged human T-cell gamma chain.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 83:2619-2623(1986).  
 CC -!- MISCELLANEOUS: THIS IS AN ABERRANTLY REARRANGED NUCLEOTIDE  
 CC SEQUENCE. TWO BASES ARE DELETED AT NUCLEOTIDE POSITION 357 IN THIS  
 CC CLONE. HOWEVER, THE PROTEIN SEQUENCE CAN BE DEDUCED FROM THE MRNA  
 CC SEQUENCE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X04038; CAA27672.1; ALT\_SEQ.  
 CC PIR; A02017; RWHUGV.  
 CC InterPro; IPR003006; -.  
 CC Pfam; PF00047; Ig; 1.  
 CC T-cell; Receptor; Glycoprotein; Signal.  
 CC KW SIGNAL 1 20  
 CC FT CHAIN 21 136 T-CELL RECEPTOR GAMMA CHAIN V REGION PT-  
 CC FT GAMMA-1/2.  
 CC FT DOMAIN 21 117 V SEGMENT.  
 CC FT DOMAIN 118 136 J SEGMENT.  
 CC FT DISULFID 41 113 POTENTIAL.  
 CC FT CARBOHYD 106 106 N-LINKED (GLCNAC..) (POTENTIAL).  
 CC SEQUENCE 136 AA; 15676 MW; AE5BB4A31DF10BA0 CRC64;  
 CC  
 CC Query Match 58.6%; Score 34; DB 1; Length 136;  
 CC Best Local Similarity 45.5%; Pred. No. 13;  
 CC Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 CC  
 CC QY 1 SSDVPCDAPLT 11  
 CC |::: || |:  
 CC Db 36 SAEITCDLFTV 46  
 CC  
 CC RESULT 8  
 CC ETFB\_BACSU STANDARD; PRT; 257 AA.  
 CC ID ETFB\_BACSU  
 CC AC P94550;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 CC DE ELECTRON TRANSFER FLAVOPROTEIN BETA-SUBUNIT (BETA-ETF) (ELECTRON  
 CC TRANSFER FLAVOPROTEIN SMALL SUBUNIT) (ETFS).  
 CC GN ETFB.  
 CC OS Bacillus subtilis.  
 CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC OC Bacillus/Staphylococcus group; Bacillus.  
 CC OX NCBI\_TaxID=1423;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=168;  
 CC RX MEDLINE=97124191; PubMed=8969504;  
 CC RA Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,  
 CC Sanders J., Emmerson P.T., Harwood C.R.;  
 CC "The dnaB-pheA (256 degrees-240 degrees) region of the Bacillus  
 CC subtilis chromosome containing genes responsible for stress  
 CC responses, the utilization of plant cell walls and primary  
 CC metabolism.";  
 CC Microbiology 142:3067-3078(1996).  
 CC -!- FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC  
 CC ELECTRON ACCEPTOR FOR OTHER DEHYDROGENASES. IT TRANSFERS THE  
 CC ELECTRONS TO THE MAIN RESPIRATORY CHAIN VIA ETF-UBIQUINONE  
 CC OXIDOREDUCTASE (ETF DEHYDROGENASE) (BY SIMILARITY).  
 CC -!- COFACTOR: CONTAINS ONE MOLECULE OF FAD PER DIMER (BY SIMILARITY).  
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.  
 CC -!- SIMILARITY: BELONGS TO THE ETF BETA-SUBUNIT / FIXA FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; 275208; CAA99574.1; -.  
 CC EMBL; Z99118; CAB14813.1; -.  
 CC HSP; P38117; IEFV.  
 CC Subtilist; BG11935; etfb.  
 CC InterPro; IPR000049; -.



DR Pfam; PF01012; ETF\_beta; 1.  
 DR PROSITE; PS01065; ETF\_BETA; 1.  
 KW Electron transport; Flavoprotein; FAD.  
 SQ SEQUENCE 257 AA; 28517 MW; 3BFC897C4121E295 CRC64;

Query Match 58.6%; Score 34; DB 1; Length 257;  
 Best Local Similarity 55.6%; Pred. No. 25;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 DVPCDATALT 11  
 :||| :|  
 Db 136 DIPCTTIT 144

RESULT 9  
 Y816\_SYNV3  
 ID Y816\_SYNV3 STANDARD; PRT; 371 AA.  
 AC P74041;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE HYPOTHETICAL 41.5 KDA PROTEIN SLL0816.  
 GN SLL0816.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -!- SIMILARITY: BELONGS TO THE GFO/IDH/MOCA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D90911; BAA18116.1; -.  
 DR InterPro; IPR000683; -.  
 DR Pfam; PF01408; GFO\_IDH\_Moca; 1.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 371 AA; 41489 MW; BDFE020D4D84BF6 CRC64;

Query Match 58.6%; Score 34; DB 1; Length 371;  
 Best Local Similarity 45.5%; Pred. No. 36;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSDVPCDATALT 11  
 :||| :|  
 Db 238 ANDVPCQINIT 248

RESULT 10  
 ARRC\_XENLA  
 ID ARRC\_XENLA STANDARD; PRT; 387 AA.  
 AC P51403;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ARRESTIN-C (CONE ARRESTIN).

OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=95237390; PubMed=7720881;  
 RA Craft C.M., Whitmore D.H.;  
 RT "The arrestin superfamily: cone arrestins are a fourth family.";  
 RL FEBS Lett. 362:247-255(1995).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN AN AS YET UNDEFINED RETINA-SPECIFIC  
 CC SIGNAL TRANSDUCTION. COULD BINDS TO PHOTOACTIVATED-PHOSPHORYLATED  
 CC RED/GREEN OPSINS.  
 CC -!- TISSUE SPECIFICITY: RETINA AND PINEAL GLAND.  
 CC -!- SIMILARITY: BELONGS TO THE ARRESTIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L40463; AAC42225.1; -.  
 DR InterPro; IPR000698; -.  
 DR Pfam; PF00339; arrestin; 1.  
 DR PRINTS; PRO0309; ARRESTIN.  
 DR PROSITE; PS00295; ARRESTINS; 1.  
 KW Sensory transduction; Vision.  
 SQ SEQUENCE 387 AA; 43045 MW; 04396764A0264266 CRC64;

Query Match 58.6%; Score 34; DB 1; Length 387;  
 Best Local Similarity 50.0%; Pred. No. 38;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSDVPCDATALT 10  
 :||| :|  
 Db 119 TIDLPCSRTL 128

RESULT 11  
 CAPP\_AMAHP  
 ID CAPP\_AMAHP STANDARD; PRT; 964 AA.  
 AC Q43299;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE (EC 4.1.1.31) (PEPCASE).  
 OS Amaranthus hypochondriacus (Prince's feather).  
 OC Eukaryota; Viridiplantae; Embryophyta; tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;  
 OC Caryophyllales; Amaranthaceae; Amaranthus.  
 OX NCBI\_TaxID=28502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Leaf;  
 RA Rydzik E., Berry J.;  
 RT "The C4 phosphoenolpyruvate carboxylase (PEPCase) from grain  
 RT Amaranth.";  
 RL (In) Plant Gene Register PCR95-135.  
 CC -!- FUNCTION: TO FORM OXALACETATE, A FOUR-CARBON DICARBOXYLIC ACID  
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.  
 CC -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALACETATE = H(2)O +  
 CC PHOSPHOENOLPYRUVATE + CO(2).  
 CC -!- ENZYME REGULATION: BY LIGHT-REVERSIBLE PHOSPHORYLATION (BY  
 CC SIMILARITY).  
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE. THIS ISOZYME IS INVOLVED IN  
 CC C4 PHOTOSYNTHESIS.  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z68125; CAA92209.1; -;  
 DR EMBL; L49175; AAB18633.1; -;  
 DR InterPro; IPR001449; -;  
 DR Pfam; PF00311; PEPCase; 1.  
 DR PRINTS; PR00150; PEPCARBLXASE.  
 DR PROSITE; PS00393; PEPCASE\_2; 1.  
 DR PROSITE; PS00781; PEPCASE\_1; 1.  
 DR Lyase; Carbon dioxide fixation; Allosteric enzyme; Multigene family;  
 KW Tricarboxylic acid cycle; Phosphorylation; Photosynthesis.  
 FT MOD\_RES 11 11 PHOSPHORYLATION (BY SIMILARITY).  
 FT ACT\_SITE 172 172 BY SIMILARITY.  
 FT ACT\_SITE 600 600 BY SIMILARITY.  
 FT ACT\_SITE 964 964 BY SIMILARITY.  
 SQ SEQUENCE 964 AA; 109480 MW; 3D49C2DE8BDE11B3 CRC64;  
 -----  
 Query Match 58.6%; Score 34; DB 1; Length 964;  
 Best Local Similarity 63.6%; Pred. No. 95;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 -----  
 Qy 1 SSDVPCDATLT 11  
 Db 391 SSDVPEESTFT 401  
 ||||| :|||  
 391 SSDVPEESTFT 401  
 -----  
 RESULT 12  
 POLG\_MVEV STANDARD; PRT; 1780 AA.  
 ID POLG\_MVEV  
 AC P05769;  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX  
 DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL  
 DE PROTEINS NS1, NS2A, AND NS2B; HELICASE (NS3)] (FRAGMENT).  
 OS Murray valley encephalitis virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OC NCBI\_TaxID=11079;  
 OX [1]  
 RW SEQUENCE FROM N.A.  
 RX MEDLINE=86200215; PubMed=3009829;  
 RA Dalgarno L., Trent D.W., Strauss J.H., Rice C.M.;  
 RT "Partial nucleotide sequence of the Murray valley encephalitis virus  
 RT genome. Comparison of the encoded polypeptides with yellow fever  
 RT virus structural and non-structural proteins.";  
 RL J. Mol. Biol. 187:309-323(1986).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X03467; CAA27184.1; -;  
 DR PIR; A24635; GNVWV.  
 DR HSP; P14336; LSVB.  
 DR MEROPS; S07.001; -;  
 DR InterPro; IPR000069; -;  
 DR InterPro; IPR000336; -;  
 DR InterPro; IPR000487; -;  
 DR InterPro; IPR000752; -;  
 DR InterPro; IPR001122; -;  
 DR InterPro; IPR001157; -;  
 DR InterPro; IPR001850; -;  
 DR InterPro; IPR002535; -;  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF00949; Flavi\_helicase; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 KW Polyprotein; Glycoprotein; Core protein; Coat protein;  
 KW Envelope protein; Helicase; ATP-binding; Transmembrane;  
 KW Nonstructural protein.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 121  
 FT PROPEP 122 217  
 FT CHAIN 218 292  
 FT CHAIN 293 793  
 FT CHAIN 794 1207  
 FT CHAIN 1208 1372  
 FT CHAIN 1373 1503  
 FT CHAIN 1504 >1780  
 FT TRANSMEM 44 60  
 FT TRANSMEM 112 128  
 FT TRANSMEM 278 294  
 FT TRANSMEM 773 789  
 FT TRANSMEM 1178 1194  
 FT TRANSMEM 1219 1235  
 FT TRANSMEM 1250 1266  
 FT TRANSMEM 1312 1328  
 FT TRANSMEM 1378 1394  
 FT TRANSMEM 1401 1417  
 FT TRANSMEM 1476 1492  
 FT DISULFID 295 322  
 FT DISULFID 352 408  
 FT DISULFID 366 397  
 FT DISULFID 384 413  
 FT DISULFID 482 580  
 FT DISULFID 597 628  
 FT CARBOHYD 73 73  
 FT CARBOHYD 140 140  
 FT CARBOHYD 446 446  
 FT CARBOHYD 923 923  
 FT CARBOHYD 968 968  
 FT CARBOHYD 1000 1000  
 FT NON\_TER 1780 1780  
 SQ SEQUENCE 1780 AA; 194866 MW; 0D6AA7FF0FB706DE CRC64;  
 -----  
 Query Match 58.6%; Score 34; DB 1; Length 1780;  
 Best Local Similarity 77.8%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 -----  
 Qy 3 DVPCDATLT 11  
 Db 1762 DVMCHATLT 1770  
 ||| | |||  
 1762 DVMCHATLT 1770  
 -----  
 RESULT 13  
 POLG\_WNV STANDARD;  
 ID POLG\_WNV  
 AC P06935;



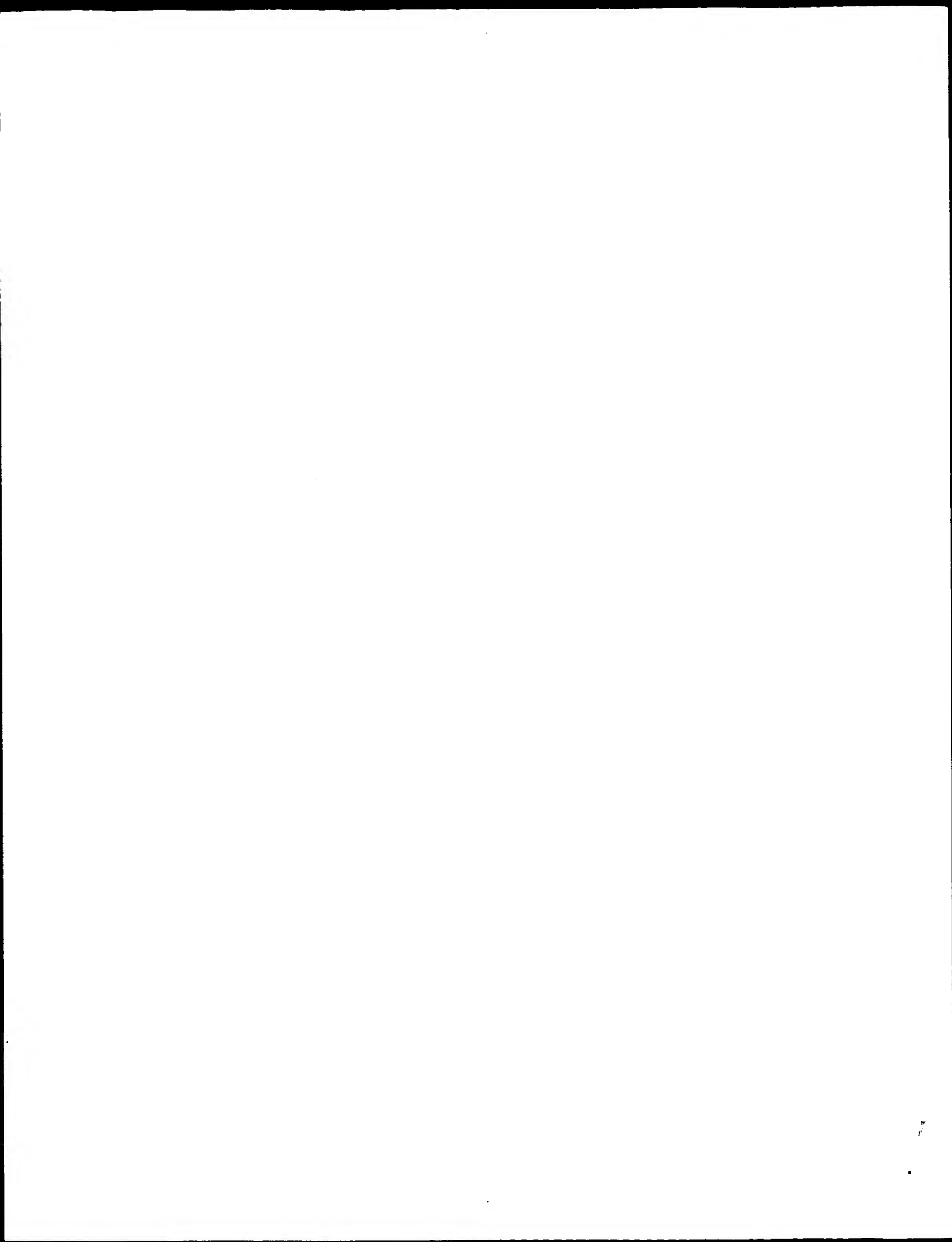


DR InterPro; IPR001528; -  
 DR InterPro; IPR001850; -  
 DR InterPro; IPR002535; -  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF00949; Flavi\_helicase; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 127  
 FT PROPEP 128 219  
 FT CHAIN 220 294  
 FT CHAIN 295 794  
 FT CHAIN 795 1206  
 FT CHAIN 1207 1373  
 FT CHAIN 1374 1504  
 FT CHAIN 1505 2123  
 FT CHAIN 2124 2412  
 FT CHAIN 2413 2527  
 FT CHAIN 2528 3432  
 FT NP\_BIND 1698 1705  
 FT SITE 1789 1792  
 FT TRANSMEM 44 60  
 FT TRANSMEM 112 127  
 FT TRANSMEM 280 294  
 FT TRANSMEM 774 790  
 FT DISULFID 297 324  
 FT DISULFID 354 410  
 FT DISULFID 368 399  
 FT DISULFID 386 415  
 FT DISULFID 484 581  
 FT DISULFID 598 629  
 FT CARBOHYD 142 142  
 FT CARBOHYD 448 448  
 FT CARBOHYD 924 924  
 FT CARBOHYD 1001 1001  
 FT CARBOHYD 2463 2463  
 FT CARBOHYD 2491 2491  
 SQ SEQUENCE 3432 AA; 380176 MW; AEB863E0C00C4674 CRC64;

Query Match 58.6%; Score 34; DB 1; Length 3432;  
 Best Local Similarity 77.8%; Pred. No. 3 4e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 DVFCDATLT 11  
 DB 1763 DVMCHATLT 1771

Search completed: May 10, 2001, 10:15:26  
 Job time: 255 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:11:45 ; Search time 33.08 Seconds  
(without alignments)  
22.852 Million cell updates/sec

Title: US-09-202-305-20  
Perfect score: 58  
Sequence: 1 SSDVPCDALT 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_67.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	100.0	267	1	RWMS8
2	58	100.0	269	2	T-cell receptor al
3	58	100.0	267	1	T-cell receptor al
4	44	75.9	220	2	T-cell receptor al
5	44	75.9	264	2	T-cell receptor al
6	44	75.9	273	2	T-cell receptor al
7	38.5	66.4	388	2	hypothetical prote
8	37	63.8	328	2	inhibin alpha-subu
9	37	63.8	406	2	hypothetical prote
10	36	62.1	139	2	hypothetical prote
11	36	62.1	290	2	organic hydroperox
12	36	62.1	486	2	phosphatidylserine
13	36	62.1	960	2	phosphoenolpyruvat
14	36	62.1	967	2	phosphoenolpyruvat
15	35	60.3	153	2	probable lipoprote
16	35	60.3	225	2	probable serine pr
17	35	60.3	472	2	hypothetical prote
18	35	60.3	633	2	polypeptide N-acet
19	35	60.3	960	1	phosphoenolpyruvat
20	35	60.3	1007	2	hypothetical prote
21	34	58.6	112	2	hypothetical prote
22	34	58.6	134	2	hypothetical prote
23	34	58.6	136	1	T-cell receptor ga
24	34	58.6	188	2	T-cell receptor ga
25	34	58.6	254	2	hypothetical prote
26	34	58.6	257	2	probable transfer
27	34	58.6	323	2	electron transfer
28	34	58.6	327	2	T-cell receptor ga
29	34	58.6	352	2	hypothetical prote
					cellulose 1,4-beta

30	34	58.6	365	2	T25843
31	34	58.6	371	2	S75555
32	34	58.6	387	2	I51404
33	34	58.6	469	2	T46929
34	34	58.6	469	2	T46930
35	34	58.6	613	2	S15468
36	34	58.6	951	2	T00260
37	34	58.6	1029	2	T30351
38	34	58.6	1353	2	T00249
39	34	58.6	1691	2	T15564
40	34	58.6	3430	1	GNWVWV
41	34	58.6	3432	1	GNWVJE
42	34	58.6	3432	1	GNWVJS
43	34	58.6	3433	1	GNWVKV
44	34	58.6	3434	1	GNWVKV
45	34	58.6	15281	2	S41309

hypothetical prote  
hypothetical prote  
cone arrestin - Af  
hypothetical prote  
hypothetical prote  
complement C3b/C4b  
hypothetical prote  
mucin-like protein  
ichl protein - ink  
hypothetical prote  
genome polyprotein  
genome polyprotein  
genome polyprotein  
genome polyprotein  
genome polyprotein  
cyclosporin synthe

#### ALIGNMENTS

RESULT 1

RWMS8

T-cell receptor alpha chain - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-May-1986 #sequence-revision 01-Dec-2000 #text\_change 01-Dec-2000

C:Accession: S25117; A93345; A93344; B26492; A02132

R:Austrup, F.; Kodellja, V.; Kucharzik, T.; Klsch, E.

submitted to the EMBL Data Library, July 1992

A:Description: Characterization of idiotypic-specific I-Ed-restricted T suppressor lymphocytes.

A:Reference number: S25117

A:Accession: S25117

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-267 <AUS>

A:Cross-references: EMBL:X67127

R:Saito, H.; Kranz, D.M.; Takagaki, Y.; Hayday, A.C.; Eisen, H.N.; Tonegawa, S.

Nature 312, 36-40, 1984

A:Title: A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes.

A:Reference number: A93345; MUID:85036635

A:Accession: A93345

A:Molecule type: mRNA

A:Residues: 130,'Y',132-267 <SAI>

A:Experimental source: strain BALB.B, clone PHDS58, cytotoxic T lymphocyte

R:Chien, Y.; Becker, D.M.; Lindsten, T.; Okamura, M.; Cohen, D.I.; Davis, M.M.

Nature 312, 31-35, 1984

A:Title: A third type of murine T-cell receptor gene.

A:Reference number: A93344; MUID:85036634

A:Accession: A93344

A:Molecule type: mRNA

A:Residues: 130,'N',132-267 <CHI>

A:Experimental source: hybridoma 2B4

R:Imai, K.; Kanno, M.; Kimoto, H.; Shigemoto, K.; Yamamoto, S.; Taniguchi, M.

Proc. Natl. Acad. Sci. U.S.A. 83, 8708-8712, 1986

A:Title: Sequence and expression of transcripts of the T-cell antigen receptor alpha-

A:Reference number: A94140; MUID:87041521

A:Accession: B26492

A:Molecule type: mRNA

A:Residues: 131-267 <IMA>

A:Experimental source: hybridoma 34S-281

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: glycoprotein; heterotetramer; receptor; T-cell; T-cell receptor; transmem

F:36-109/Domain: immunoglobulin homology <IMM>

F:130-241/Domain: C region #status predicted <CON>

F:242-262/Domain: transmembrane #status predicted <TMN>

F:263-267/Domain: intracellular #status predicted <INT>

F:197,211,240/Binding site: carbohydrate (Asn) #status predicted

Query Match

Best Local Similarity 100.0%; Score 58; DB 1; Length 267;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSDVPCDATLT 11

|||||

Db 216 SSDVPCDATLT 226

RESULT 2

S03715

T-cell receptor alpha chain precursor (F5) - mouse

C:Species: Mus musculus (house mouse)

C>Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 23-Jul-1999

C:Accession: S03715

R:Palmer, M.S.; Bentley, A.; Gould, K.; Townsend, A.R.M.

Nucleic Acids Res. 17, 2353, 1989

A>Title: The T cell receptor from an influenza-A specific murine CTL clone.

A:Reference number: S03715; MUID:89202046

A:Accession: S03715

A:Molecule type: mRNA

A:Residues: 1-269 <PAL>

A:Cross-references: EMBL:X14387; NID:g54509; PIDN:CAA32562.1; PID:g54510

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: T-cell receptor

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-263/Product: T-cell receptor alpha chain #status predicted <MAT>

Query Match

Best Local Similarity 100.0%; Score 58; DB 2; Length 269;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSDVPCDATLT 11

|||||

Db 218 SSDVPCDATLT 228

RESULT 3

JC2337

T-cell receptor alpha-chain - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 02-Jul-1998

C:Accession: JC2337

R:Stangel, M.; Toyka, K.V.; Hartung, H.P.; Giegerich, G.

Biochem. Biophys. Res. Commun. 202, 1280-1284, 1994

A>Title: Bacterial expression of a soluble T-cell receptor alpha chain.

A:Reference number: JC2337; MUID:94338338

A:Accession: JC2337

A:Molecule type: mRNA

A:Residues: 1-220 <STA>

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: receptor

F:85-90/Region: V-element

F:91-129/Region: J-element

F:130-136/Region: C-element

Query Match

Best Local Similarity 75.9%; Score 44; DB 2; Length 220;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSDVPCDATLT 10

|||||

Db 197 TSDVPCDAKL 206

RESULT 4

F27579

T-cell receptor alpha chain precursor V and C regions (TRA29) - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 08-Mar-1989 #sequence\_revision 30-Jun-1991 #text\_change 23-Jul-1999

C:Accession: F27579

R:Morris, M.; Barclay, A.N.; Williams, A.F.

Immunogenetics 27, 174-179, 1988

A>Title: Analysis of T cell receptor beta chains in rat thymus, and rat C-alpha and C-beta

A:Reference number: A27578; MUID:88113841

A:Accession: F27579

A:Molecule type: mRNA

A:Residues: 1-264 <MOR>

A:Cross-references: EMBL:M18853; NID:g207163; PIDN:AAA42207.1; PID:g207164

A>Note: the authors translated the codon CTC for residue 5 as val

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: T-cell receptor

Query Match

Best Local Similarity 75.9%; Score 44; DB 2; Length 264;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSDVPCDATLT 10

|||||

Db 213 TSDVPCDAKL 222

RESULT 5

PL0063

T-cell receptor alpha chain precursor V-J-C region - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999

C:Accession: PL0063

R:Burns, F.R.; Li, X.; Shen, N.; Offner, H.; Chou, Y.K.; Vandenbark, A.A.; Heber-Katz

J. Exp. Med. 169, 27-39, 1989

A>Title: Both rat and mouse T cell receptors specific for the encephalitogenic determ

complex and encephalitogenic determinants being recognized are different.

A:Reference number: PL0063; MUID:89080488

A:Accession: PL0063

A:Molecule type: mRNA

A:Residues: 1-273 <BUR>

A:Cross-references: GB:X14318; NID:g57751; PIDN:CAA32498.1; PID:g762996

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: heterodimer; T-cell receptor

F:1-27/Domain: signal sequence (fragment) #status predicted <SIG>

F:28-273/Product: T-cell receptor alpha chain #status predicted <MAT>

F:28-118/Domain: V region #status predicted <VRE>

F:126-138/Domain: J region #status predicted <JRE>

F:139-273/Domain: C region #status predicted <CRE>

Query Match

Best Local Similarity 75.9%; Score 44; DB 2; Length 273;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSDVPCDATLT 10

|||||

Db 222 TSDVPCDAKL 231

RESULT 6

T47530

hypothetical protein F16L2.110 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47530

R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24468

A:Accession: T47530

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-388 <JOR>

A:Cross-references: EMBL:AL162459

A:Experimental source: cultivar Columbia; BAC clone F16L2

C:Genetics:

A:Map position: 3

A:Introns: 189/3; 208/2; 272/3

A>Note: F16L2.110



```

Query Match      66.4%; Score 38.5; DB 2; Length 388;
Best Local Similarity 90.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 SSDVPCDRTL 10
Db 274 SSD-PCDRTL 282

RESULT 7
I51215
inhibin alpha-subunit precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 17-Mar-2000
C:Accession: I51215
R:Wang, S.Y.; Johnson, P.A.
Biol. Reprod. 49, 453-458, 1993
A:Title: Complementary deoxyribonucleic acid cloning and sequence analysis of the alpha-
A:Reference number: I51215; MUID:94002740
A:Accession: I51215
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-328 <WAN>
A:Cross-references: GB:S65963; NID:g430815; PID:g430816
C:Superfamily: inhibin

Query Match      63.8%; Score 37; DB 2; Length 328;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSDVPCDRTL 9
Db 83 STDVPCPT 91

RESULT 8
T30650
hypothetical protein 48L - Molluscum contagiosum virus 1
N:Alternate names: MC048L
C:Species: Molluscum contagiosum virus 1
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jul-2000
C:Accession: T30650
R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.
Science 273, 813-816, 1996
A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re
A:Reference number: 220876; MUID:96325459
A:Accession: T30650
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-406 <SEN>
A:Cross-references: EMBL:U60315; NID:g1491943; PIDN:AAC55176.1; PID:g1491991
C:Genetics:
A:Note: MC048L
C:Superfamily: vaccinia virus I6 protein

Query Match      63.8%; Score 37; DB 2; Length 406;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 DVPCDRTL 11
Db 286 DLPCDRTL 294

RESULT 9
B75489
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: B75489
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

```

```

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: B75489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <WHI>
A:Cross-references: GB:AE001925; GB:AE000513; NID:g6458383; PIDN:AAF10267.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0685
A:Map position: 1

```

```

Query Match      63.8%; Score 37; DB 2; Length 713;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 DVPCDRTL 11
Db 224 DVPCDRTL 232

RESULT 10
E75346
organic hydroperoxide resistance protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75346
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: E75346
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-139 <WHI>
A:Cross-references: GB:AE002025; GB:AE000513; NID:g6459627; PIDN:AAF11408.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1857
A:Map position: 1
C:Superfamily: hypothetical protein yk1A

```

```

Query Match      62.1%; Score 36; DB 2; Length 139;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 DVPCDRTL 11
Db 72 DVPCDRTL 80

RESULT 11
I64051
phosphatidylserine decarboxylase (EC 4.1.1.65) precursor - Haemophilus influenzae (st
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Nov-2000
C:Accession: I64051
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Godayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: I64051
A:Status: nucleic acid sequence not shown; translation not shown

```

A:Molecule type: DNA  
A:Residues: 1-290 <TIGR>  
A:Cross-references: GB:U32701; GB:L42023; NID:g3212182; PIDN:AAC21829.1; PID:g1573117; T  
C:Superfamily: Escherichia coli phosphatidylserine decarboxylase  
C:Keywords: blocked amino end; carbon-carbon lyase; carboxy-lyase; phospholipid biosynth  
F;256-257/cleavage site: Gly-Ser (autolytic) #status predicted  
F;257/Modified site: pyruvic acid (Ser) (in mature form) #status predicted

Query Match 62.1%; Score 36; DB 2; Length 290;  
Best Local Similarity 85.7%; Pred. No. 33;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VPCDATT 10  
:|||||  
Db 154 MPCDATT 160

RESULT 12  
PC2169  
phosphoenolpyruvate carboxykinase (pyrophosphate) (EC 4.1.1.38), PE105 - rape (fragments  
C:Species: Brassica napus (rape)  
C:Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 17-Mar-1999  
C:Accession: PC2169  
R:Yanai, Y.; Okumura, S.; Shimada, H.  
Biosci. Biotechnol. Biochem. 58, 950-953, 1994  
A:Title: Structure of Brassica napus phosphoenolpyruvate carboxylase genes: Missing intr  
A:Reference number: PC2167; MUID:94289869  
A:Accession: PC2169  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-486 <YAN>  
A:Cross-references: DBJ:D26379  
C:Superfamily: phosphoenolpyruvate carboxylase  
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 62.1%; Score 36; DB 2; Length 486;  
Best Local Similarity 60.0%; Pred. No. 55;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SDVPCDATT 11  
||||:|  
Db 140 SDIPAEATFT 149

RESULT 13  
SI8240  
phosphoenolpyruvate carboxylase (EC 4.1.1.31) - sorghum  
C:Species: Sorghum bicolor (sorghum)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999  
C:Accession: SI8240  
R:Lepiniec, L.; Sauti, S.; Keryer, E.; Amiet, V.; Vidal, J.; Gadai, P.; Cretin, C.  
Plant Mol. Biol. 17, 1077-1079, 1991  
A:Title: Complete nucleotide sequence of one member of the Sorghum phosphoenolpyruvate c  
A:Reference number: SI8240; MUID:92032766  
A:Accession: SI8240  
A:Molecule type: DNA  
A:Residues: 1-960 <LEP>  
A:Cross-references: EMBL:X59925; NID:g22592; PIDN:CAA42549.1; PID:g22593  
A:Note: the authors translated the codon TTG for residue 395 as Phe, AAT for residue 696  
C:Genetics:  
A:Introns: 51/3; 182/3; 210/3; 284/3; 319/3; 348/3; 400/3; 733/3; 862/3  
C:Superfamily: phosphoenolpyruvate carboxylase  
C:Keywords: carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase

Query Match 62.1%; Score 36; DB 2; Length 960;  
Best Local Similarity 70.0%; Pred. No. 1.le+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SDVPCDATT 11  
||||:|

Db 387 SDIPEEATLT 396

## RESULT 14

JH0667

phosphoenolpyruvate carboxylase (EC 4.1.1.31) C3-form - maize

C:Species: Zea mays (maize)

C:Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 22-Jun-1999

C:Accession: JH0667

R:Kawamura, T.; Shigesada, K.; Toh, H.; Okumura, S.; Yanagisawa, S.; Izui, K.

J. Biochem. 112, 147-154, 1992

A:Title: Molecular evolution of phosphoenolpyruvate carboxylase for C4 photosynthesis  
tic function.

A:Reference number: JH0667; MUID:93054411

A:Accession: JH0667

A:Molecule type: mRNA

A:Residues: 1-967 &lt;KAW&gt;

A:Cross-references: GB:X61489; NID:g429148; PIDN:CAA43709.1; PID:g429149

A:Experimental source: root

C:Comment: This enzyme catalyzes the carboxylation of phosphoenolpyruvate to form oxa

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 62.1%; Score 36; DB 2; Length 967;

Best Local Similarity 70.0%; Pred. No. 1.le+02;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SDVPCDATT 11

||||:|

Db 394 SDIPEEATLT 403

## RESULT 15

T37219

probable lipoprotein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T37219

R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z21615

A:Accession: T37219

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-153 &lt;OLI&gt;

A:Cross-references: EMBL:AL031514; PIDN:CAA20597.1; GSPDB:GN00070; SCOEDB:SC2H4.03

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC2H4.03

Query Match 60.3%; Score 35; DB 2; Length 153;

Best Local Similarity 70.0%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 SDVPCDATT 11

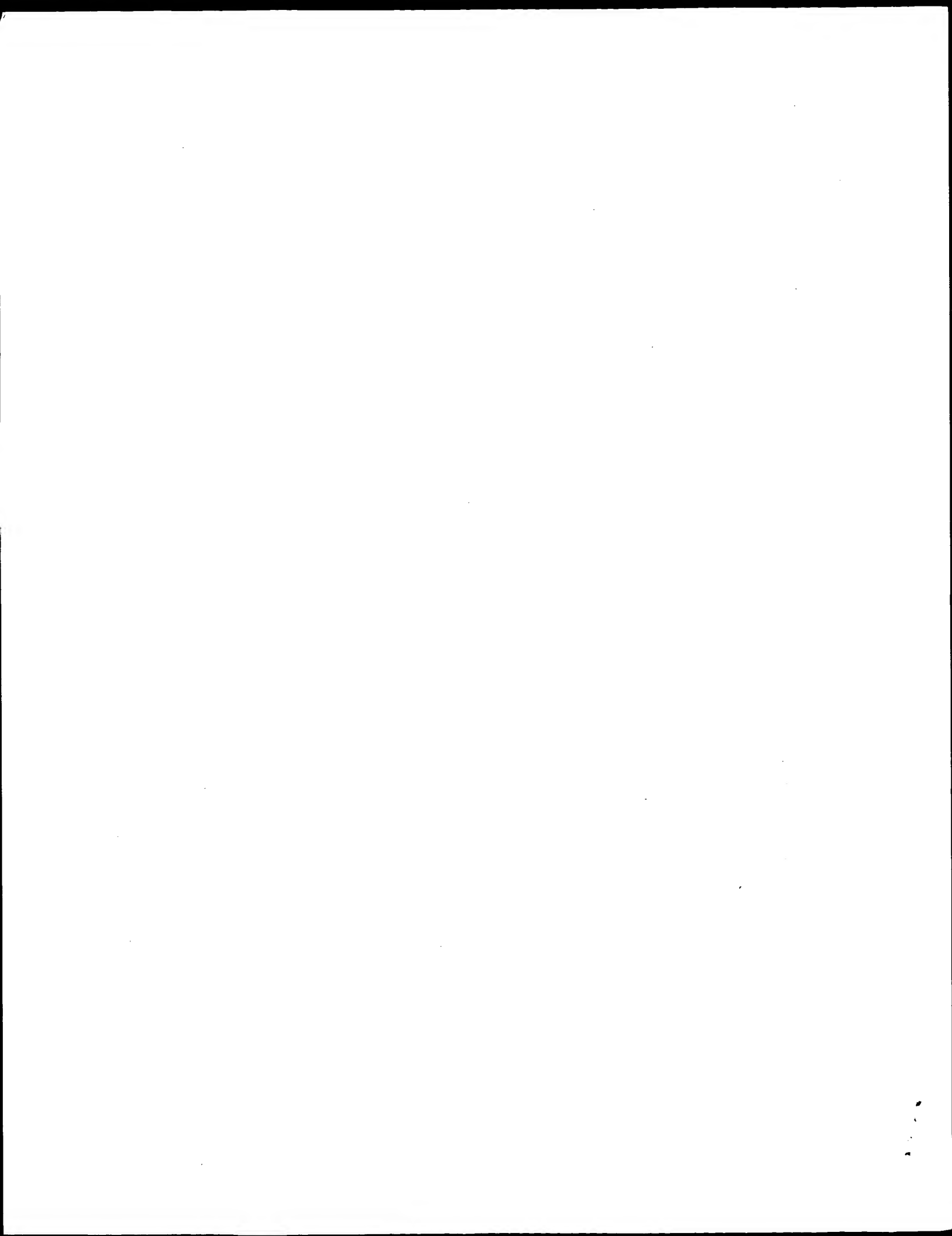
||||:|

Db 116 SDEPCRLT 125

Search completed: May 10, 2001, 10:11:47

Job time: 132 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:10:35 ; Search time 51.82 Seconds  
(without alignments)  
12.134 Million cell updates/sec

Title: US-09-202-305-20  
Perfect score: 58  
Sequence: 1 SSDVFCDAFLT 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0401.\*

```
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	58	100.0	11	W42371	TCR-alpha chain de
2	58	100.0	109	W36107	Mouse T-cell recep
3	58	100.0	215	W7287	T-cell receptor al
4	58	100.0	266	Y05727	Mouse Al T cell re
5	58	100.0	267	W04300	Murine T-cell rece
6	58	100.0	268	P60237	Sequence of the al
7	58	100.0	268	W7288	T-cell receptor al
8	58	100.0	268	W36110	Mouse T-cell recep
9	58	100.0	270	P50256	Sequence of T-cell
10	58	100.0	272	W36111	Mouse T-cell recep
11	58	100.0	273	Y05405	Killer T-cell recep

12	58	100.0	274	20	Y05404	Killer T-cell rece
13	38	65.5	124	21	G54964	Arabidopsis thalia
14	38	65.5	170	15	R65969	Glioblastoma-deriv
15	37	63.8	1817	21	B42230	Human OREFX ORF1994
16	36	62.1	153	17	R98466	Alpha chain of sol
17	36	62.1	911	15	R52568	ppCase gene in la
18	35	60.3	77	19	Y20843	Human neurofilamen
19	35	60.3	387	18	W18663	Fragmented human N
20	35	60.3	607	21	W36172	Arabidopsis thalia
21	35	60.3	607	21	Y67411	Arabidopsis pyruva
22	34	58.6	323	14	R32906	TCR gamma protein
23	34	58.6	324	10	P91885	Deduced sequence o
24	34	58.6	329	20	W86010	Mouse CAAX process
25	34	58.6	329	21	Y55810	Mouse RCE1 (farnes
26	34	58.6	951	21	B21253	Human metalloprote
27	34	58.6	15281	15	R44929	T. niveum cyclospo
28	33	56.9	10	15	R58761	Foki 58 kD tryptic
29	33	56.9	10	16	R72803	Foki tryptic pepti
30	33	56.9	10	16	R66488	Foki restriction e
31	33	56.9	147	20	Y73831	Human prostate tum
32	33	56.9	234	19	W53973	Pseudomonas fluore
33	33	56.9	328	21	Y84620	Amino acid sequenc
34	33	56.9	519	21	Y44247	Human cell signal
35	33	56.9	570	17	R96300	Foki insertion mut
36	33	56.9	573	17	R96299	Foki insertion mut
37	33	56.9	577	15	R58759	Foki amino acid se
38	33	56.9	579	16	R72810	Foki. Flavobacter
39	33	56.9	581	17	R88594	Foki insertion mut
40	33	56.9	581	17	R88593	Foki insertion mut
41	33	56.9	583	11	R06978	Modified enzyme fr
42	33	56.9	585	17	R88595	Foki insertion mut
43	33	56.9	589	12	R14325	Rat epithelin prec
44	33	56.9	589	17	R88596	Foki insertion mut
45	33	56.9	595	17	R88599	Foki insertion mut

#### ALIGNMENTS

RESULT	1
W42371	ID W42371 standard; peptide: 11 AA.
XX	AC W42371;
XX	XX
DT	11-JUN-1998 (first entry)
XX	XX
DE	TCR-alpha chain derived peptide.
XX	XX
KW	Peptide: TCR-alpha chain; T-cell receptor; inflammation; infection;
KW	autoimmune disease; inflammatory bowel disease; psoriasis;
KW	acquired immune deficiency syndrome; allergy.
XX	XX
OS	Synthetic.
OS	Homo sapiens.
XX	XX
PN	W09747644-Al.
XX	XX
PD	18-DEC-1997.
XX	XX
PF	11-JUN-1997; 97WO-AU00367.
XX	XX
PR	11-JUN-1996; 96AU-0000394.
PR	11-JUN-1996; 96AU-0000389.
PR	11-JUN-1996; 96AU-0000390.
PR	11-JUN-1996; 96AU-0000391.
PR	11-JUN-1996; 96AU-0000392.
PR	11-JUN-1996; 96AU-0000393.
XX	XX
PA	(NSVD-) NORTHERN SYDNEY AREA HEALTH SERVICE.
XX	XX
PI	Manolios N;
XX	XX

DR WPI: 1998-052238/05.  
 XX New peptide(s) that inhibit the T cell receptor - used to treat  
 PT inflammation, auto-immune disease, allergy etc. and to deliver  
 XX conjugated therapeutic agents to cells  
 XX  
 PS Claim 12; Page 32; 58pp; English.  
 XX  
 CC The present peptide sequence is derived from the T-cell receptor  
 CC (TCR)-alpha chain. This peptide and others (see W42357-W42371) act to  
 CC inhibit TCR function probably by interfering with TCR assembly. They are  
 CC claimed to be useful in treating disorders in which T cells are involved  
 CC or recruited, e.g. allergy, autoimmune diseases, inflammatory bowel  
 CC disease, psoriasis, infections (including acquired immune deficiency  
 CC syndrome), etc. They are also claimed to be able to deliver conjugated  
 CC therapeutic agents to cells.  
 XX  
 SQ Sequence 11 AA;  
 Query Match 100.0%; Score 58; DB 19; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.00016; Mismatches 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 0;  
 QY 1 SSDVPCDALT 11  
 Db 1 ssdvpcdatlt 11  
 RESULT 2  
 W36107 W36107 standard; protein; 109 AA.  
 AC W36107;  
 XX 19-MAY-1998 (first entry)  
 DT Mouse T-cell receptor alpha-chain constant region.  
 DE Mouse: T-cell receptor; alpha-chain constant region; antigen-specific;  
 KW immunosuppressant; humoral; cell mediated immune response; allergy;  
 KW hypersensitivity; autoimmune reaction; transplant rejection.  
 XX Mus sp.  
 OS W09743411-AL.  
 PN 20-NOV-1997.  
 XX 09-MAY-1997; 97WO-JP01565.  
 PF 29-MAY-1996; 96JP-0135572.  
 PR 10-MAY-1996; 96JP-0116101.  
 XX (KIRI ) KIRIN BEER KK.  
 PA Honma N, Mikayama T, Yuyama N;  
 PI WPI: 1998-008880/01.  
 DR Immunosuppressant peptide containing T-cell receptor alpha-chain  
 PT sequence - are not antigen-specific and do not induce antibody  
 PT production  
 PS Claim 2; Page 35; 63pp; Japanese.  
 XX The present sequence represents mouse T-cell receptor alpha-chain  
 CC constant region. The protein is an immunosuppressant which is not  
 CC antigen-specific and suppresses both humoral and cell-mediated immune  
 CC reactions. It can be used for treatment and/or prevention of delayed  
 CC hypersensitivity reactions, allergies and autoimmune reactions, and  
 CC inhibition of transplant rejection. The protein does not induce the  
 CC formation of antibodies against them to any significant extent.

XX Sequence 109 AA;  
 SQ Query Match 100.0%; Score 58; DB 19; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 0.0018; Mismatches 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 0;  
 QY 1 SSDVPCDALT 11  
 Db 86 ssdvpcdatlt 96  
 RESULT 3  
 R77287 R77287 standard; Protein; 215 AA.  
 ID R77287;  
 AC R77287;  
 XX 24-NOV-1995 (first entry)  
 DT T-cell receptor alpha chain (TCR alpha) from hybridoma A1.1.  
 DE T-cell receptor alpha chain: TCR alpha; hyperimmune;  
 XX immunodeficiency; Cd4+; helper T cell; hybridoma A1.1.  
 KW Synthetic.  
 OS W09516462-A.  
 PN 22-JUN-1995.  
 XX 13-DEC-1994; 94WO-US14542.  
 PF 13-DEC-1993; 93US-0165496.  
 PR (KIRI ) KIRIN BEER KK.  
 XX (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.  
 PA Bissonnette R, Fotedar A, Green D, Ishii Y, Mikayama T;  
 XX WPI: 1995-231357/30.  
 DR N-PSDB; Q91361.  
 XX Antigen specific immune modulation using T cell receptor alpha chain -  
 PT for treating allergy, cancer, auto-immune disease etc.  
 PS Example; Figure 14; 119pp; English.  
 XX TCR alpha cDNA of CD4+, helper T cell hybridoma A1.1 cells was  
 CC cloned from cDNA library using C alpha probes. mRNA was isolated  
 CC from 10(9) cells and recovered. The first strand cDNA was  
 CC synthesised using an oligo-dT primer and reverse transcriptase and  
 CC the second strand using DNA polymerase I and RNase H. The ds cDNA  
 CC was ligated to EcoRI linkers, digested with EcoRI, and cloned into  
 CC lambda-gt10. The phage was packaged and screened using  
 CC radiolabelled C alpha probes. The DNA from positive clones were  
 CC prep'd. for sequencing and the complete nt. sequence of A1.1 TCR  
 CC alpha cDNA is shown in Q91361. The deduced AA sequence is in  
 CC R77287.  
 XX Sequence 215 AA;  
 SQ Query Match 100.0%; Score 58; DB 16; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 0.0037; Mismatches 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 0;  
 QY 1 SSDVPCDALT 11  
 Db 192 ssdvpcdatlt 202

RESULT 4

Y05727  
 XX Y05727 standard; Protein; 266 AA.  
 AC Y05727;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE Mouse Al T cell receptor alpha chain.  
 XX  
 KW T cell receptor alpha chain; mouse; transgenic animal;  
 KW animal model; immunological tolerance; graft rejection;  
 KW tissue grafting; neonatal intolerance; transplantation antigen;  
 KW H-Y antigen.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09916867-Al.  
 XX  
 PD 08-APR-1999.  
 XX  
 PF 30-SEP-1998; 98WO-GB02965.  
 XX  
 PR 01-OCT-1997; 97GB-0020888.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Cobbold SP, Waldmann H, Zelenika D;  
 XX  
 DR WPI: 1999-255090/21.  
 DR N-PSDB; X25357.  
 XX  
 PT Transgenic non-human mammal having only CD4 positive T cells  
 PT specific for at least one transplantation antigen, useful for  
 PT studying immunological tolerance  
 XX  
 PS Disclosure; Page 32-33; 41pp; English.  
 XX

The present sequence represents the T cell receptor (TCR) alpha chain from the Al CD4+ T cell clone isolated from CRA/Ca mice. The Al clone recognises the minor histocompatibility antigen H-Y. present in male, but absent in female, mice. CDNA (see X25357) encoding the Al TCR alpha chain was obtained by RT-PCR (see also X25350-51), and was used in the construction of Al(M) transgenic mice. The invention relates to a transgenic animal model comprising TCR alpha and beta chains. It provides a genetically modified non-human mammal having a population of CD4 positive T cells specific for one or a limited number of selected antigens, including at least transplantation antigen capable of rejecting a tissue transplant containing the transplantation antigen and, if applicable, the other selected antigens. The animal has TCR genes which encode a TCR specific for the transplantation antigen. The animal is useful for studying immunological tolerance, especially the mechanisms of tolerance to, and the rejection of, tissue grafts, and in pregnancy. The animals are also useful for testing agents for biological activity in promoting or reducing immunological tolerance.

XX Sequence 266 AA;

Query Match 100.0%; Score 58; DB 20; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSDVPCDATT 11  
 Db 215 ssdvpdatl 225  
 |||||

RESULT 5

W04300  
 ID W04300 standard; Protein; 267 AA.

XX W04300;  
 AC  
 XX 02-JAN-1997 (first entry)  
 DT  
 DE Murine T-cell receptor alpha chain.  
 XX  
 KW Antigen-specific glycosylation inhibiting factor; AgGIF; lymphokine;  
 KW T-cell receptor; TCR; immunosuppressant; immunosuppressive;  
 KW hypersensitivity; graft rejection; autoimmune disease.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..20  
 FT /label= Leader\_peptide  
 FT Region 21..109  
 FT /label= V-alpha\_region  
 FT Region 110..130  
 FT /label= J-alpha\_region  
 FT Region 131..267  
 FT /label= C-alpha\_region  
 XX  
 PN W09631617-Al.  
 XX  
 PD 10-OCT-1996.  
 XX  
 PF 25-SEP-1995; 95WO-US12309.  
 XX  
 PR 04-APR-1995; 95US-0416336.  
 XX  
 PA (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.  
 XX  
 PI Ishii Y, Ishizaka K;  
 XX  
 DR WPI: 1996-465037/46.  
 DR N-PSDB; T38389.  
 XX  
 PT Recombinant antigen-specific glycosylation-inhibiting factor -  
 PT useful as immunosuppressant  
 XX  
 PS Example 3; Fig 2A; 72pp; English.  
 XX

A cDNA clone (T38389) coding for the T-cell receptor (TCR) alpha chain (W04300) was isolated from T-cell hybridoma 231F1 cells (ovalbumin-specific). TCR alpha plays a role in the formation of antigen-specific glycosylation inhibiting factor (AgGIF) chains which directly bind to an antigen and suppress the immune response generated against that antigen; AgGIF is an expression product of the TCR alpha chain gene (see also T38389). A polynucleotide encoding AgGIF can be obtd. from activated suppressor T-cells using a probe based on non-specific GIF cDNA (see also T38390-91) and a probe based on TCR alpha cDNA, and used to direct expression of AgGIF in host cells.

XX Sequence 267 AA;

Query Match 100.0%; Score 58; DB 17; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSDVPCDATT 11  
 Db 216 ssdvpdatl 226  
 |||||

RESULT 6

P60237  
 ID P60237 standard; Protein; 268 AA.

XX P60237;  
 AC

19-AUG-1991 (first entry)

Sequence of the alpha subunit of a clonally diversified integral membrane protein encoded by clone PHDS58.

T-cell clone specific antibody: cytotoxic T lymphocyte; immunoassay; drug delivery.

Mus musculus.

Key Location/Qualifiers

Peptide 1..20 /label= signal

Region 21..118 /label= variable

Region 119..131 /label= joining

Region 132..241 /label= constant

Region 242..263 /label= transmembrane

Region 264..268 /label= cytoplasmic

Binding-site 90..92 /label= potential N-glycosylation site

Binding-site 198..200 /label= see above

Binding-site 212..214 /label= see above

Binding-site 241..243 /label= see above

Disulfide-bond 22..110 /label= see above

Disulfide-bond 154..204 /note= "inter-chain"

EP180878-A.

14-MAY-1986.

26-OCT-1985; 85EP-0113658.

31-OCT-1984; 84US-0666988.

13-JUN-1984; 84US-0620122.

01-MAR-1984; 84US-0585333.

22-OCT-1984; 84US-0663809.

27-JUL-1989; 89US-0385897.

(MASI ) MASSACHUSETTS INST TECH.

(STRD ) LELAND STANFORD JR UNIV.

Saito H, Kranz DM, Eisen HN, Tonegawa S;

WPI; 1986-126342/20.

N-PSDB; N60166.

New hetero-dimeric t-lymphocyte receptor and its sub-units - useful in prodn. of T-cell clone specific antibodies for immunoassays, isolation procedures, specific delivery of bound drugs etc.

Example; Fig 4a; 50pp; English.

T-cell specific cDNA clones were isolated from an alloreactive CTL clone 2C of Balb/c mouse origin and specific for the D end of the Balb/C H-2 complex. A library of cDNA was constructed and screened with hybridisation probes. Two distinct classes of T-cell specific cDNA clones whose genes are rearranged in cytotoxic T lymphocytes (CTLs) were identified (see N60166 and N60167). A heterodimeric T lymphocyte receptor comprising an alpha and a beta subunit is claimed.

Sequence 268 AA;

Query Match 100.0%; Score 58; DB 7; Length 268;

Best Local Similarity 100.0%; Pred. No. 0.0046;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSDVPCDATT 11

DB 217 ssdvpcdatlt 227

RESULT 7

R77288

ID R77288 standard; Protein; 268 AA.

XX R77288;

XX 24-NOV-1995 (first entry)

DT T-cell receptor alpha chain (TCR alpha) from hybridoma 3B3.

DE T-cell receptor alpha chain: TCR alpha; hyperimmune;

XX immunodeficiency; Cd4+; helper T cell; hybridoma 3B3.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..20

FT /label= signal

XX WO9516462-A.

XX 22-JUN-1995.

XX 13-DEC-1994; 94WO-US14542.

XX 13-DEC-1993; 93US-0165496.

XX (KIRI ) KIRIN BEER KK.

XX (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.

XX Bissonnette R, Fotedar A, Green D, Ishii Y, Mikayama T;

XX WPI; 1995-231357/30.

XX N-PSDB; Q91362.

XX Antigen specific immune modulation using T cell receptor alpha chain - for treating allergy, cancer, auto-immune disease etc.

XX Example; Figure 15; 119pp; English.

XX TCR alpha cDNA of 3B3 cells was cloned by PCR. mRNA was isolated from 3B3 cells. cDNA was generated. cDNAs were ligated at the 5' and 3' ends using T4 ligase to construct circular DNA. Oligo primers encoding murine C alpha DNA were synthesised (see Q91363, Q91364). PCR was carried out. Amplified cDNA was subcloned into CCRI000 vector of TA cloning system. Three different TCR alpha cDNA were cloned and sequenced. Two of them were identified to be CC originated from the fusion partner cell of 3B3 hybridoma, BW5147. CC The other was confirmed not to be expressed in BW5147 which CC indicated that this TCR alpha originated from PLA2-specific T CC cells. Two of independent clones encoding this TCR alpha cDNA were CC isolated and their DNA sequences were confirmed to be identical. CC The DNA sequence of this 3B3 derived TCR alpha cDNA is shown in CC Q91362 and its deduced AA sequence is in R77288.

XX Sequence 268 AA;

Query Match 100.0%; Score 58; DB 16; Length 268;

Best Local Similarity 100.0%; Pred. No. 0.0046;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 SSDVPCDATALT 11  
 DB 217 ssdvpdatlt 227

RESULT 8  
 W36110  
 ID W36110 standard; Protein: 268 AA.  
 XX AC W36110;  
 XX DT 19-MAY-1998 (first entry)  
 XX DE Mouse T-cell receptor alpha-chain protein from cell line 3B3.  
 XX KW Mouse; T-cell receptor; alpha-chain constant region; antigen-specific;  
 KW immunosuppressant; humoral; cell mediated immune response; allergy;  
 KW hypersensitivity; autoimmune reaction; transplant rejection.  
 XX OS Mus sp.  
 XX PN W09743411-A1.  
 XX PD 20-NOV-1997.  
 XX PF 09-MAY-1997; 97WO-JP01565.  
 XX PR 29-MAY-1996; 96JP-0135572.  
 XX PR 10-MAY-1996; 96JP-0116101.  
 XX PA (KIRI ) KIRIN BEER KK.  
 XX PI Honma N, Mikayama T, Yuyama N;  
 XX DR WPI; 1998-008880/01.  
 XX DR N-PSDB; V01408.  
 XX PT Immunosuppressant peptide containing T-cell receptor alpha-chain  
 PT sequence - are not antigen-specific and do not induce antibody  
 PT production  
 XX PS Example 1; Page 37-38; 63pp; Japanese.  
 XX CC The present sequence represents mouse T-cell receptor alpha-chain  
 CC from cell line 3B3. The protein is an immunosuppressant which is not  
 CC antigen-specific and suppresses both humoral and cell-mediated immune  
 CC reactions. It can be used for treatment and/or prevention of delayed  
 CC hypersensitivity reactions, allergies and autoimmune reactions, and  
 CC inhibition of transplant rejection. The protein does not induce the  
 CC formation of antibodies against them to any significant extent.  
 XX SQ Sequence 268 AA;  
 Query Match 100.0%; Score 58; DB 19; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SSDVPCDATALT 11  
 DB 217 ssdvpdatlt 227

RESULT 9  
 P50256  
 ID P50256 standard; Protein: 270 AA.  
 XX AC P50256;  
 XX DT 07-OCT-1991 (first entry)  
 XX DE Sequence of T-cell antigen receptor alpha chain encoded by cDNA  
 DE clone Tt11.

Diagnosis; site-directed therapy.  
 Homo sapiens.  
 Key Location/Qualifiers  
 Peptide 1..20  
 Region 21..114  
 Region 115..131  
 Region 132..249  
 Region 250..265  
 Region 266..270  
 Modified-site 42..44  
 Modified-site 200..202  
 Modified-site 214..216  
 Modified-site 243..245  
 Note="possible carbohydrate attachment site"  
 Note="as above"  
 Note="as above"  
 Note="as above"  
 W08503947-A.  
 12-SEP-1985.  
 28-FEB-1985; 85WO-US00367.  
 22-OCT-1984; 84US-0663809.  
 01-MAR-1984; 84US-0585333.  
 31-OCT-1984; 84US-0666988.  
 (STRD ) LELAND STANFORD UNI.  
 Davis MM, Hedrick SM;  
 WPI; 1985-249152/40.  
 N-PSDB; N50284.  
 New DNA sequences coding for T-cell antigen receptors or  
 fragments - useful in diagnostic assays, affinity chromatography,  
 site directed therapy and diagnosis  
 Disclosure; Fig 3; 41pp; English.  
 Mammalian T-cell receptors appear to be 80-90kda heterodimers,  
 which are disulphide linked, and composed of two distinct  
 glycoproteins of about 40 to 50kd, referred to as the alpha- and  
 beta- subunits (N50284, P50256; N50280, P50252 respectively). The  
 chains may be divided up into regions associated with specific exons  
 by analogy to immunoglobulins. Genes encoding helper T-cell antigen-  
 specific receptor subunits alpha- and beta- (TH-Ag receptor, alpha-  
 or beta-subunit) were isolated. For the beta-subunit three thymus-  
 derived clones were obtd. designated 86T1, 86T3 and 86T5 (N50280-  
 N50282). The inventors claim a DNA sequence of at least about 15nt  
 present in the sequence of 86T1 (N50280) or Tt11 (N50284) joined to  
 non-wild type DNA.  
 Query Match 100.0%; Score 58; DB 6; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 0.0047;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SSDVPCDATALT 11  
 DB 219 ssdvpdatlt 229

```

RESULT 10
W36111
ID W36111 standard; Protein: 272 AA.
XX
XX AC W36111;
XX
XX 19-MAY-1998 (first entry)
XX
XX Mouse T-cell receptor alpha-chain protein from cell line B4-9.52.
XX
XX Mouse; T-cell receptor; alpha-chain constant region; antigen-specific;
XX immunosuppressant; humoral; cell mediated immune response; allergy;
XX hypersensitivity; autoimmune reaction; transplant rejection.
XX
XX Mus sp.
XX
XX W09743411-A1.
XX
XX 20-NOV-1997.
XX
XX 09-MAY-1997; 97WO-JP01565.
XX
XX 29-MAY-1996; 96JP-0135572.
XX
XX 10-MAY-1996; 96JP-0116101.
XX
XX (KIRI ) KIRIN BEER KK.
XX
XX Honma N, Mikayama T, Yuyama N;
XX
XX WPI: 1998-008880/01.
XX N-PSDB; V01417.
XX
XX Immunosuppressant peptide containing T-cell receptor alpha-chain
XX sequence - are not antigen-specific and do not induce antibody
XX production
XX
XX Example 9; Page 39-40; 63pp; Japanese.
XX
XX The present sequence represents mouse T-cell receptor alpha-chain from
XX cell line B4-9.52. The protein is an immunosuppressant which is not
XX antigen-specific and suppresses both humoral and cell-mediated immune
XX reactions. It can be used for treatment and/or prevention of delayed
XX hypersensitivity reactions, allergies and autoimmune reactions, and
XX inhibition of transplant rejection. The protein does not induce the
XX formation of antibodies against them to any significant extent.
XX
XX Sequence 272 AA;
XX
XX Query Match 100.0%; Score 58; DB 19; Length 272;
XX Best Local Similarity 100.0%; Pred. No. 0.0047;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SSDVPCDATLT 11
XX |||||||
XX Db 221 ssdvpdatlt 231
XX
XX RESULT 11
XX Y05405
XX ID Y05405 standard; Protein: 273 AA.
XX
XX AC Y05405;
XX
XX 02-JUL-1999 (first entry)
XX
XX Killer T-cell receptor protein sequence.
XX
XX Killer T-cell receptor; HIV; infection; HIV-1 IIIB; therapy.
XX
XX Mus musculus.
XX

```

```

XX
XX PN W09916885-A1.
XX
XX PD 08-APR-1999.
XX
XX PF 28-SEP-1998; 98WO-JP04345.
XX
XX PR 26-SEP-1997; 97JP-0262536.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX (SAIT/) SAITO T.
XX (TAKA/) TAKAHASHI H.
XX
XX Saito T, Takahashi H;
XX
XX WPI: 1999-255096/21.
XX N-PSDB; X36398.
XX
XX Killer T-cell receptor peptide specifically recognizing HIV-infected
XX cells
XX
XX Example 3; Page 68-70; 75pp; Japanese.
XX
XX This sequence is a killer T-cell receptor protein of the invention,
XX which recognises and damages cells infected with human immunodeficiency
XX virus (HIV), especially with HIV-1 IIIB. The receptor can be used as
XX a constituent of compositions for the treatment of HIV infection.
XX
XX Sequence 273 AA;
XX
XX Query Match 100.0%; Score 58; DB 20; Length 273;
XX Best Local Similarity 100.0%; Pred. No. 0.0047;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SSDVPCDATLT 11
XX |||||||
XX Db 222 ssdvpdatlt 232
XX
XX RESULT 12
XX Y05404
XX ID Y05404 standard; Protein: 274 AA.
XX
XX AC Y05404;
XX
XX 02-JUL-1999 (first entry)
XX
XX Killer T-cell receptor protein sequence.
XX
XX Killer T-cell receptor; HIV; infection; HIV-1 IIIB; therapy.
XX
XX Mus musculus.
XX
XX W09916885-A1.
XX
XX 08-APR-1999.
XX
XX 28-SEP-1998; 98WO-JP04345.
XX
XX 26-SEP-1997; 97JP-0262536.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX (SAIT/) SAITO T.
XX (TAKA/) TAKAHASHI H.
XX
XX Saito T, Takahashi H;
XX
XX WPI: 1999-255096/21.
XX N-PSDB; X36392.
XX
XX Killer T-cell receptor peptide specifically recognizing HIV-infected
XX cells
XX

```

XX Claim 7; Page 64-66; 75pp; Japanese.  
XX This sequence is a killer T-cell receptor protein of the invention.  
CC which recognises and damages cells infected with human immunodeficiency  
CC virus (HIV), especially with HIV-1 IIIB, the receptor can be used as  
CC a constituent of compositions for the treatment of HIV infection.  
XX  
SQ Sequence 274 AA;

Query Match 100.0%; Score 58; DB 20; Length 274;  
Best Local Similarity 100.0%; Pred. No. 0.0047;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSDVPCDATLT 11  
|||||||  
Db 223 ssdvpdatlt 233

RESULT 13  
G54964  
ID G54964 standard; Protein; 124 AA.  
XX  
AC G54964;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 70329.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.  
25-MAY-1999; 99US-0136021.  
27-MAY-1999; 99US-0136392.  
28-MAY-1999; 99US-0136782.  
01-JUN-1999; 99US-0137222.  
03-JUN-1999; 99US-0137528.  
04-JUN-1999; 99US-0137502.  
07-JUN-1999; 99US-0137724.  
08-JUN-1999; 99US-0138094.  
10-JUN-1999; 99US-0138540.  
10-JUN-1999; 99US-0138847.  
14-JUN-1999; 99US-0139119.  
16-JUN-1999; 99US-0139452.  
16-JUN-1999; 99US-0139453.  
17-JUN-1999; 99US-0139492.  
18-JUN-1999; 99US-0139454.  
18-JUN-1999; 99US-0139455.  
18-JUN-1999; 99US-0139456.  
18-JUN-1999; 99US-0139457.  
18-JUN-1999; 99US-0139458.  
18-JUN-1999; 99US-0139459.  
18-JUN-1999; 99US-0139460.  
18-JUN-1999; 99US-0139461.  
18-JUN-1999; 99US-0139462.  
18-JUN-1999; 99US-0139463.  
18-JUN-1999; 99US-0139750.  
18-JUN-1999; 99US-0139763.  
21-JUN-1999; 99US-0139817.  
22-JUN-1999; 99US-0139899.  
23-JUN-1999; 99US-0140353.  
23-JUN-1999; 99US-0140354.  
24-JUN-1999; 99US-0140695.  
28-JUN-1999; 99US-0140823.  
29-JUN-1999; 99US-0140991.  
30-JUN-1999; 99US-0141287.  
01-JUL-1999; 99US-0141842.  
01-JUL-1999; 99US-0142154.  
02-JUL-1999; 99US-0142055.  
06-JUL-1999; 99US-0142390.  
08-JUL-1999; 99US-0142803.  
09-JUL-1999; 99US-0142920.  
12-JUL-1999; 99US-0142977.  
13-JUL-1999; 99US-0143542.  
14-JUL-1999; 99US-0143624.  
15-JUL-1999; 99US-0144005.  
16-JUL-1999; 99US-0144085.  
16-JUL-1999; 99US-0144086.  
19-JUL-1999; 99US-0144325.  
19-JUL-1999; 99US-0144331.  
19-JUL-1999; 99US-0144332.  
19-JUL-1999; 99US-0144333.  
19-JUL-1999; 99US-0144334.  
19-JUL-1999; 99US-0144335.  
20-JUL-1999; 99US-0144352.  
20-JUL-1999; 99US-0144632.  
20-JUL-1999; 99US-0144884.  
21-JUL-1999; 99US-0144814.  
21-JUL-1999; 99US-0145086.  
21-JUL-1999; 99US-0145088.  
22-JUL-1999; 99US-0145085.  
22-JUL-1999; 99US-0145087.  
22-JUL-1999; 99US-0145089.  
22-JUL-1999; 99US-0145192.  
23-JUL-1999; 99US-0145145.  
23-JUL-1999; 99US-0145218.  
23-JUL-1999; 99US-0145224.  
26-JUL-1999; 99US-0145276.  
27-JUL-1999; 99US-0145913.  
27-JUL-1999; 99US-0145918.  
27-JUL-1999; 99US-0145919.  
28-JUL-1999; 99US-0145951.  
02-AUG-1999; 99US-0146386.  
02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149002.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 65.5%; Score 38; DB 21; Length 124;  
Best Local Similarity 70.0%; Pred. No. 9.8; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 2 SDVPCDALT 11  
|||||  
DB 105 sdvpcsqalt 114  
|||||

RESULT 14  
R65969  
ID R65969 standard; Protein; 170 AA.  
XX R65969;  
AC R65969;  
XX  
DT 24-JUN-1995 (first entry)  
XX  
DE Glioblastoma-derived polypeptide.  
XX  
KW Glioblastoma; therapeutic.  
XX  
OS Homo sapiens T98G.  
XX  
PN EP624646-A.  
XX  
PD 17-NOV-1994.  
XX  
PF 04-MAY-1994; 94EP-0107001.  
XX  
PR 11-MAY-1993; 93JP-0109130.  
XX  
PA (ONOV) ONO PHARM CO LTD.  
XX  
PI Hirano A, Miyamoto T, Shibayama S;  
XX  
DR WPI; 1994-350786/44.  
DR N-PSDB; Q78292.  
XX  
PT New polypeptide from human glioblastoma cell - and related  
PT nucleic acid, vectors, antibodies, etc.. useful in treatment and  
PT diagnosis of e.g. aplasia, cell proliferation, inflammation, etc.  
XX  
PS Claim 1; Page 10-11; 17pp; English.  
XX  
CC The polypeptide is produced by the human glioblastoma cell line T98G  
CC (ATCC CRL 1690) and can be used in the treatment of aplasia or  
CC abnormal proliferation of glia, neurons or hematopoietic cells,  
CC depression or enhancement of immunological or neurological  
CC activities, inflammation and tumors.  
XX  
SQ Sequence 170 AA;

Query Match 65.5%; Score 38; DB 15; Length 170;  
Best Local Similarity 62.5%; Pred. No. 14;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SDVPCDALT 9  
:|||||  
DB 158 adipcdst 165  
:|||||

RESULT 15  
B42230  
ID B42230 standard; Protein; 1817 AA.  
XX  
AC B42230;  
XX  
DT 08-FEB-2001 (first entry)

XX DE Human ORFX ORF1994 polypeptide sequence SEQ ID NO:3988.  
 XX XX  
 KW Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX XX  
 OS Homo sapiens.  
 XX XX  
 PN WO200058473-A2.  
 XX XX  
 PD 05-OCT-2000.  
 XX XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX XX  
 FA (CURA-) CURAGEN CORP.  
 XX XX  
 PI Shinkets RA, Leach M;  
 XX XX  
 DR WPI: 2000-602362/57.  
 DR N-PSDB; C76439.  
 XX XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX XX  
 PS Claim 11: Page 3156-3161; 5507pp; English.  
 XX XX  
 CC C7446 to C77606 encode the proteins given in B40237 to B43397, which  
 CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX XX  
 SQ Sequence 1817 AA;

|||||  
 Db 304 ssdlpcd 310  
 Search completed: May 10, 2001, 10:10:37  
 Job time: 62 sec

Query Match 63.8%; Score 37; DB 21; Length 1817;  
 Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SSDVPCD 7



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:10:41 ; Search time 53.33 seconds  
(without alignments)  
19,780 Million cell updates/sec

Title: US-09-202-305-6  
Perfect score: 41  
Sequence: 1 ILLKLVAGF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_15.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	78.0	50	10 Q41532	Q41532 triticum ae
2	32	78.0	166	3 Q9UR21	Q9ur21 schizosacch
3	32	78.0	369	2 Q9KAV0	Q9kav0 bacillus ha
4	32	78.0	409	2 Q9RTL7	Q9rtl7 deinococcus
5	32	78.0	467	2 Q84914	Q84914 lactobacill
6	31	75.6	297	2 P74131	P74131 synecocyst
7	31	75.6	305	10 Q96566	Q96566 hydrilla ve
8	31	75.6	364	10 Q96567	Q96567 hydrilla ve
9	31	75.6	367	10 Q9M4K3	Q9m4k3 bucegia rom
10	31	75.6	367	10 Q9M4I8	Q9m4i8 preissia qu
11	31	75.6	368	2 Q9RK10	Q9rk10 streptomyce
12	31	75.6	368	10 Q9M4J5	Q9m4j5 lunularia c
13	31	75.6	368	10 Q9M4J2	Q9m4j2 marchantia
14	31	75.6	370	10 Q9M3H4	Q9m3h4 epidendrum
15	31	75.6	375	1 Q57876	Q57876 pyrococcus
16	31	75.6	418	2 Q9PEN5	Q9pen5 xylella fas
17	31	75.6	455	4 Q9UIR0	Q9uir0 homo sapien
18	31	75.6	669	10 Q04914	Q04914 psilocum nu
19	31	75.6	964	10 Q9SCB2	Q9scb2 lycopersico

20	31	75.6	1983	5 Q9NEP3	Q9nep3 caenorhabdi
21	30	73.2	145	2 Q9X033	Q9x033 thermotoga
22	30	73.2	198	4 Q9NWV0	Q9nwv0 homo sapien
23	30	73.2	203	2 Q9RVP3	Q9rvp3 deinococcus
24	30	73.2	206	2 Q06690	Q06690 treponema p
25	30	73.2	222	10 Q9MAP3	Q9map3 arabidopsis
26	30	73.2	223	5 Q9NDX7	Q9ndx7 theileria s
27	30	73.2	262	10 Q92W14	Q92w14 arabidopsis
28	30	73.2	293	2 P74949	P74949 vibrio orda
29	30	73.2	367	11 Q9RI81	Q9ri81 mus musculu
30	30	73.2	369	10 Q49641	Q49641 arabidopsis
31	30	73.2	369	10 Q9M4I5	Q9m4i5 scapania ne
32	30	73.2	375	1 Q9V2E0	Q9v2e0 pyrococcus
33	30	73.2	379	10 Q9M4J6	Q9m4j6 jungermanni
34	30	73.2	389	2 Q25184	Q25184 helicobacte
35	30	73.2	408	2 Q26036	Q26036 helicobacte
36	30	73.2	408	2 Q9ZJB4	Q9zjb4 helicobacte
37	30	73.2	468	3 P87276	P87276 saccharomyc
38	30	73.2	470	3 P87277	P87277 saccharomyc
39	30	73.2	470	3 P87278	P87278 saccharomyc
40	30	73.2	470	3 Q06595	Q06595 saccharomyc
41	30	73.2	470	3 Q06719	Q06719 saccharomyc
42	30	73.2	470	3 Q9UVH5	Q9uvh5 saccharomyc
43	30	73.2	478	1 Q9YE79	Q9ye79 aeropyrum p
44	30	73.2	556	2 Q9KWN3	Q9kwn3 sphingomona
45	30	73.2	572	2 Q9PBK4	Q9pbk4 xylella fas

## ALIGNMENTS

RESULT	1
Q41532	
ID	Q41532 PRELIMINARY; PRT; 50 AA.
AC	Q41532;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE	ALPHA-AMYLASE 2/53 (FRAGMENT).
GN	AMY1
OS	Triticum aestivum (Wheat).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
OX	NCBI_TaxID=4565;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. CHINESE SPRING;
RX	MEDLINE=89181522; PubMed=2467183;
RA	Huttly A.K., Martienssen R.A., Baulcombe D.C.;
RT	"Sequence heterogeneity and differential expression of the alpha-Amy2
RT	gene family in wheat."
RL	Mol. Gen. Genet. 214:232-240(1988).
DR	EMBL; X13577; CAA31914.1; -
DR	HSSP; P04063; IAVA.
DR	MENDEL; 10961; Triae;Amy1;10961.
FT	NON_TER 50
SQ	SEQUENCE 50 AA; 5588 MW; P9159A9A131226D2 CRC64;

Query Match 78.0%; Score 32; DB 10; Length 50;  
Best Local Similarity 77.8%; Pred. No. 17;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
Db 12 LLLLLVAGF 20

RESULT 2  
Q9UR21 PRELIMINARY; PRT; 166 AA.  
ID Q9UR21  
AC Q9UR21;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE HYPOTHETICAL 19.5 KDA PROTEIN.  
 GN SPAC806.10 OR SPAC24B11.14.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
 OC Schizosaccharomycetaceae; Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H;  
 RA Wedler H., Duesterhoeft A., Wood V., Rajandream M.A., Barrell B.G.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H;  
 RA Odell C., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL117212; CAB61651.1; -;  
 DR EMBL; Z67757; CAB54976.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 166 AA; 19503 MW; 5B4FD3D595ADAC86 CRC64;

Query Match 78.0%; Score 32; DB 3; Length 166;  
 Best Local Similarity 75.0%; Pred. No. 57;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLKVAGF 9  
 :|:|:|:|  
 Db 65 ILVRVAGF 72

RESULT 3  
 Q9KAV0 ID Q9KAV0 PRELIMINARY; PRT; 369 AA.  
 AC Q9KAV0;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE BH2186 PROTEIN.  
 GN BH2186.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RA Takami H., Nakasone K., Takaki Y.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP001514; BAB05905.1; -;  
 SQ SEQUENCE 369 AA; 42179 MW; 3964AFC7AD07C632 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 369;  
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ILLKVAGF 9  
 :|:|:|:|  
 Db 8 ILLVAVAGF 16

RESULT 4  
 Q9RTL7 ID Q9RTL7 PRELIMINARY; PRT; 409 AA.  
 AC Q9RTL7;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE FOSMIDOMYCIN RESISTANCE PROTEIN, PUTATIVE.  
 GN DR1743.

OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Ueberback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome Sequence of the Radioresistant Bacterium Deinococcus  
 radiodurans R1."  
 RL Science 286:1571-1577(1999).  
 DR EMBL; AE002016; AAF11300.1; -;  
 DR TIGR; DR1743; -;  
 DR INTERPRO; IPR001066; -;  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
 SQ SEQUENCE 409 AA; 42175 MW; BDBE53B2DF5970B2 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 409;  
 Best Local Similarity 77.8%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ILLKVAGF 9  
 :|:|:|:|  
 Db 127 ILLVAVAGF 135

RESULT 5  
 O84914 ID O84914 PRELIMINARY; PRT; 467 AA.  
 AC O84914;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CARNOSINASE.  
 GN PEPV.  
 OS Lactobacillus helveticus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=1587;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CNR32;  
 RA Shao W., Yuksel G.U., Parkin K.L., Steele J.L.;  
 RT "Biochemical and molecular characterization of pepv from Lactobacillus  
 helveticus CNR32."  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF012085; AAC24967.1; -;  
 DR MEROPS; M20.004; -;  
 DR INTERPRO; IPR001261; -;  
 DR INTERPRO; IPR002933; -;  
 DR PFAM; PF01546; Peptidase\_M20; 1.  
 DR PROSITE; PS00758; ARGE\_DAPE\_CPG2.1; 1.  
 DR PROSITE; PS00759; ARGE\_DAPE\_CPG2.2; 1.  
 SQ SEQUENCE 467 AA; 51487 MW; FD7531A80D3E3E8B CRC64;

Query Match 78.0%; Score 32; DB 2; Length 467;  
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ILLKVAGF 9  
 :|:|:|:|  
 Db 130 MLLKKEAGF 138

RESULT 6



P74131  
ID P74131 PRELIMINARY; PRT; 297 AA.  
AC P74131;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE HYPOTHETICAL 33.5 KDA PROTEIN.  
GN SLR1980.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirose M., Sugiyama M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D90912; BAA18217.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 297 AA; 33479 MW; B251968942F479C2 CRC64;

Query Match 75.6%; Score 31; DB 2; Length 297;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ILLKLVAGF 9  
:|:|:|:|:|  
Db 116 LLLLVIAGF 124

RESULT 7  
ID Q96566 PRELIMINARY; PRT; 305 AA.  
AC Q96566;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE PHOSPHOENOLPYRUVATE CARBOXYLASE ISOFORM 1 (EC 4.1.1.31) (FRAGMENT).  
GN PPC1.  
OS Eukarya; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Hydrocharitaceae; Hydrilla.  
OX NCBI\_TaxID=51024;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ROYLE;  
RA Magnin N., Reiskind J.B., Bowes G.;  
RT "Identification of Phosphoenolpyruvate carboxylase isoforms from an  
RT aquatic monocot with inducible C4-type photosynthesis.";  
RL Plant Physiol. 111:72-72(1996).  
CC -1- CATALYTIC ACTIVITY: PHOSPHATE + OXALOACETATE = H(2)O +  
CC PHOSPHOENOLPYRUVATE + CO(2).  
DR EMBL; U65226; AAB08698.1; -  
DR HSP; P00864; IFIY.  
DR MENDEL; 15963; Hydve; Pp1:15962.  
DR INTERPRO; IPR001449; -  
DR PFAM; PF00311; PEPcase; 1.  
DR PRINTS; PR00150; PEPcase; 1.  
DR KW Pyruvate; Lyase.  
FT NON\_TER 1  
SQ SEQUENCE 305 AA; 34531 MW; 37C9BC63309E4196 CRC64;

Query Match 75.6%; Score 31; DB 10; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8  
:|:|:|:|:|  
Db 269 LLLKVAG 275

RESULT 8  
ID Q96567 PRELIMINARY; PRT; 364 AA.  
AC Q96567;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE PHOSPHOENOLPYRUVATE CARBOXYLASE ISOFORM 2 (EC 4.1.1.31) (FRAGMENT).  
GN PPC1.  
OS Hydrilla verticillata (hydrilla).  
OC Eukarya; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Hydrocharitaceae; Hydrilla.  
OX NCBI\_TaxID=51024;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Magnin N., Reiskind J.B., Bowes G.;  
RT "Identification of Phosphoenolpyruvate carboxylase isoforms from an  
RT aquatic monocot with inducible C4-type photosynthesis.";  
RL Plant Physiol. 111:72-72(1996).  
CC -1- CATALYTIC ACTIVITY: PHOSPHATE + OXALOACETATE = H(2)O +  
CC PHOSPHOENOLPYRUVATE + CO(2).  
DR EMBL; U65227; AAB08697.1; -  
DR HSP; P00864; IFIY.  
DR MENDEL; 15963; Hydve; Pp1:15963.  
DR INTERPRO; IPR001449; -  
DR PFAM; PF00311; PEPcase; 1.  
DR PRINTS; PR00150; PEPcase; 1.  
DR KW Pyruvate; Lyase.  
FT NON\_TER 1  
SQ SEQUENCE 364 AA; 41002 MW; 9A02A7AEF2BFC985 CRC64;

Query Match 75.6%; Score 31; DB 10; Length 364;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8  
:|:|:|:|:|  
Db 269 LLLKVAG 275

RESULT 9  
ID Q9M4K3 PRELIMINARY; PRT; 367 AA.  
AC Q9M4K3;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE PHOSPHOENOLPYRUVATE CARBOXYLASE (EC 4.1.1.31) (FRAGMENT).  
GN PPC.  
OS Bucegia romana.  
OC Eukarya; Viridiplantae; Embryophyta; Marchantiophyta; Marchantiales;  
OC Marchantiaceae; Bucegia.  
OX NCBI\_TaxID=53027;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Genig H.H., Reck U.;  
RT "Molecular phylogeny of Bryophytes and Pteridophytes based on partial  
RT amino acid sequences of the phosphoenolpyruvate carboxylase (PEPC).";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ231280; CAB90612.1; -  
DR KW Pyruvate; Lyase.  
FT NON\_TER 1  
SQ SEQUENCE 367 AA; 41144 MW; 62D9D5068265F440 CRC64;

Query Match 75.6%; Score 31; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8  
 DB 269 LLLKVAG 275

RESULT 10  
 Q9M4I8 PRELIMINARY; PRT; 367 AA.  
 AC Q9M4I8;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE (EC 4.1.1.31) (FRAGMENT).  
 GN PPC.  
 OS Preissia quadrata.  
 OC Eukaryota; Viridiplantae; Embryophyta; Marchantiophyta; Marchantiales;  
 OC Marchantiaceae; Preissia.  
 OX NCBI\_TaxID=53029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gehrig H.H., Reck U.;  
 RT "Molecular phylogeny of Bryophytes and Pteridophytes based on partial  
 RT amino acid sequences of the phosphoenolpyruvate carboxylase (PEPC).";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ231297; CAB90681.1;  
 KW Pyruvate; Lyase.  
 FT NON\_TER 1  
 SQ SEQUENCE 367 AA; 41297 MW; 0363CA23C3FD27D0 CRC64;

Query Match 75.6%; Score 31; DB 10; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8  
 DB 269 LLLKVAG 275

RESULT 11  
 Q9RK10 PRELIMINARY; PRT; 368 AA.  
 AC Q9RK10;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE PUTATIVE IRON-SIDOPHORE PERMEASE TRANSMEMBRANE PROTEIN.  
 GN SCF34.15C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL109974; CAB5326.1;  
 DR INTERPRO; IPR000522;  
 DR PFAM; PF01032; PfecCD\_family; 1.  
 KW Transmembrane.  
 SQ SEQUENCE 368 AA; 37547 MW; 8FBC80D0A5F88D31 CRC64;

Query Match 75.6%; Score 31; DB 2; Length 368;  
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLKVAG 9  
 DB 47 LLLAVAG 54

RESULT 12  
 Q9M4J5 PRELIMINARY; PRT; 368 AA.  
 AC Q9M4J5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE (EC 4.1.1.31) (FRAGMENT).  
 GN PPC.  
 OS Lunularia cruciata.  
 OC Eukaryota; Viridiplantae; Embryophyta; Marchantiophyta; Marchantiales;  
 OC Lunulariaceae; Lunularia.  
 OX NCBI\_TaxID=56931;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gehrig H.H., Reck U.;  
 RT "Molecular phylogeny of Bryophytes and Pteridophytes based on partial  
 RT amino acid sequences of the phosphoenolpyruvate carboxylase (PEPC).";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ231289; CAB90662.1;  
 KW Pyruvate; Lyase.  
 FT NON\_TER 1  
 SQ SEQUENCE 368 AA; 41469 MW; 4548F82FBB9A27DC CRC64;

Query Match 75.6%; Score 31; DB 10; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8  
 DB 269 LLLKVAG 275

RESULT 13  
 Q9M4J2 PRELIMINARY; PRT; 368 AA.  
 AC Q9M4J2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE (EC 4.1.1.31) (FRAGMENT).  
 GN PPC.  
 OS Marchantia calcarata.  
 OC Eukaryota; Viridiplantae; Embryophyta; Marchantiophyta; Marchantiales;  
 OC Marchantiaceae; Marchantia.  
 OX NCBI\_TaxID=80892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gehrig H.H., Reck U.;  
 RT "Molecular phylogeny of Bryophytes and Pteridophytes based on partial  
 RT amino acid sequences of the phosphoenolpyruvate carboxylase (PEPC).";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ231292; CAB90665.1;  
 KW Pyruvate; Lyase.  
 FT NON\_TER 1

SQ SEQUENCE 368 AA; 41384 MW; D8C97D7077479FAC CRC64;

Query Match 75.6%; Score 31; DB 10; Length 368;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8  
| | | | |  
DB 269 LLLKVAG 275

## RESULT 14

Q9M3H4 ID Q9M3H4 PRELIMINARY; PRT; 370 AA.  
AC Q9M3H4;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE PHOSPHOENOLPYRUVATE CARBOXYLASE (EC 4.1.1.31) (FRAGMENT).  
GN PPC.  
OS Epidendrum stamfordianum.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Asparagales; Orchidaceae; Epidendrum.  
OX NCBI\_TaxID=126619;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAF;  
RA Gehrig H.H.;  
RT "New PEPC sequences."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ400987; CAB92916.1; -  
KW Pyruvate; Lyase.  
FT NON\_TER 1  
SQ SEQUENCE 370 AA; 41813 MW; 892342C6797ADD5B CRC64;

Query Match 75.6%; Score 31; DB 10; Length 370;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8  
| | | | |  
DB 269 LLLKVAG 275

## RESULT 15

O57876 ID O57876 PRELIMINARY; PRT; 375 AA.  
AC O57876;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE 375AA LONG HYPOTHETICAL NA(+)/H(+) ANTIporter.  
GN PH0136.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3."  
RL DNA Res. 5:55-76(1998).  
DR EMBL; AP000001; BAA29205.1; -  
SQ SEQUENCE 375 AA; 40970 MW; 385686747E575FDA CRC64;

Query Match 75.6%; Score 31; DB 1; Length 375;  
Best Local Similarity 75.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 9  
| | | | |  
DB 176 ILLEVAGF 183

Search completed: May 10, 2001, 10:14:59  
Job time: 258 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:11:11 ; Search time 20.13 seconds  
(without alignments)  
15.315 Million cell updates/sec

Title: US-09-202-305-6  
Perfect score: 41  
Sequence: 1 ILLKLVAGF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	41	100.0	138	1	TCA_MOUSE
2	41	100.0	142	1	TCA_HUMAN
3	34	82.9	481	1	YEAV_ECOLI
4	32	78.0	470	1	PEPV_LACDL
5	32	78.0	473	1	MAGR_YEAST
6	31	75.6	298	1	PORR_METJA
7	30	73.2	295	1	SNAI_BOVIN
8	30	73.2	295	1	SNAI_HUMAN
9	30	73.2	295	1	SNAI_RAT
10	30	73.2	293	1	YKCF_BACSU
11	30	73.2	299	1	YKAM_BACSU
12	30	73.2	407	1	OPPB_MYCGE
13	30	73.2	452	1	YEDV_ECOLI
14	29	70.7	334	1	YE96_MYCTU
15	29	70.7	437	1	TIG_NEIMA
16	29	70.7	437	1	TIG_NEIMB
17	29	70.7	809	1	NAH2_RABIT
18	29	70.7	812	1	NAH2_HUMAN
19	29	70.7	850	1	RNI2_YEAST
20	28	68.3	153	1	YEHR_ECOLI
21	28	68.3	179	1	YCXY_CVAPA
22	28	68.3	191	1	API2_PETMA
23	28	68.3	235	1	PHU2_THEMA
24	28	68.3	242	1	CCMC_RHOCA
25	28	68.3	267	1	TRUA_CHLTR
26	28	68.3	283	1	CHER_BORBU
27	28	68.3	366	1	TRUB_TREPA
28	28	68.3	384	1	RN_DROME
29	28	68.3	391	1	YIDY_ECOLI
30	28	68.3	430	1	AST2_YEAST
31	28	68.3	440	1	DGTP_SYNY3
32	28	68.3	465	1	YFF2_YEAST
33	28	68.3	475	1	YMT4_YEAST

34	28	68.3	520	1	RP54_RHIET
35	28	68.3	524	1	VGIG_RABVS
36	28	68.3	527	1	AOPA_BOVIN
37	28	68.3	545	1	ICAI_RAT
38	28	68.3	570	1	DUD_ECOLI
39	28	68.3	677	1	WHIT_LUCCU
40	28	68.3	698	1	COX1_TETPY
41	28	68.3	960	1	CAP2_MESCR
42	28	68.3	1407	1	CYAA_DICDI
43	28	68.3	1584	1	YJ9G_YEAST
44	28	68.3	3421	1	TEGU_HSVB
45	27	65.9	122	1	PA23_TRIGA

P49989 rhizobium e  
P16288 rabies viru  
P21398 bos taurus  
Q00238 rattus norv  
P06149 escherichia  
Q05360 lucilia cup  
P11947 tetrahymena  
P16097 mesembryant  
Q03100 dictyosteli  
P27170 saccharomyc  
P28955 equine herp  
P81480 trimeresuru

## ALIGNMENTS

RESULT 1  
TCA\_MOUSE  
ID TCA\_MOUSE STANDARD; PRT; 138 AA.  
AC P01849;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 21-JUL-1986 (Rel. 01, Last annotation update)  
DE T-CELL RECEPTOR ALPHA CHAIN C REGION.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (CLONE PY14).  
RC STRAIN=BALB.B;  
RX MEDLINE=85036635; PubMed=6208487;  
RA Saito H., Kranz D.M., Takagaki Y., Hayday A.C., Eisen H.N.,  
RA Tonegawa S.;  
RT "A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes";  
RL Nature 312:36-40(1984).  
RN [2]  
RP SEQUENCE FROM N.A. (HYBRIDOMA 2B4).  
RX MEDLINE=85036634; PubMed=6548551;  
RA Chien Y., Becker D.M., Lindsten T., Okamura M., Cohen D.I.,  
RA Davis M.M.;  
RT "A third type of murine T-cell receptor gene";  
RL Nature 312:31-35(1984).  
CC -!- MISCELLANEOUS: CLONE PHDS58 WAS ISOLATED FROM A CYTOTOXIC T LYMPHOCYTE.  
DR PIR: A02132; RWMSC8.  
KW T-cell; Receptor; Transmembrane; Glycoprotein.  
FT NON\_TER 1 1  
FT DOMAIN 1 112 C SEGMENT.  
FT TRANSMEM 113 133  
FT DOMAIN 134 138 CYTOPLASMIC TAIL.  
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .).  
FT CARBOHYD 82 82 N-LINKED (GLCNAC. .).  
FT CARBOHYD 111 111 N-LINKED (GLCNAC. .).  
FT VARIANT 2 2 Y -> N (IN HYBRIDOMA 2B4).  
SQ SEQUENCE 138 AA; 15494 MW; 984B77953AA80444 CRC64;

Query Match 100.0%; Score 41; DB 1; Length 138;  
Best Local Similarity 100.0%; Pred. No. 0.08;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILLKLVAGF 9  
Db 119 ILLKLVAGF 127

RESULT 2  
TCA\_HUMAN  
ID TCA\_HUMAN STANDARD; PRT; 142 AA.  
AC P01848;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE T-CELL RECEPTOR ALPHA CHAIN C REGION.  
 GN TRAC.  
 GN Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (CLONE PV14).  
 RX MEDLINE=85216514; PubMed=3873654;  
 RA Yanagi Y., Chan A., Chin B., Minden M., Mak T.W.;  
 RT "Analysis of cDNA clones specific for human T cells and the alpha and  
 RT beta chains of the T-cell receptor heterodimer from a human T-cell  
 RT line.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3430-3434(1985).  
 CC -1- MISCELLANEOUS: THIS CLONE WAS ISOLATED FROM A HUMAN LEUKEMIC  
 CC T-CELL LINE, JURKAT.  
 CC -1- MISCELLANEOUS: THIS ALPHA CHAIN C REGION SHOWS SEQUENCE HOMOLOGY  
 CC TO ITS BETA CHAIN COUNTERPART.  
 CC -1- MISCELLANEOUS: THE GENE CORRESPONDING TO THIS MRNA IS REARRANGED  
 CC SPECIFICALLY IN T-CELLS; ITS ORGANIZATION IS SIMILAR TO AN IG  
 CC GENE, WITH V, D, J, AND C REGIONS.  
 CC PIR; A02131; RWHUAC.  
 KW T-cell; Receptor; Transmembrane; MHC.  
 FT NON\_TER 1 1  
 FT DOMAIN 1 117 C REGION.  
 FT TRANSMEM 118 137  
 FT DOMAIN 138 142 CYTOPLASMIC TAIL.  
 SQ SEQUENCE 142 AA; 15928 MW; D46B44171D5784EE CRC64;  
  
 Query Match 100.0%; Score 41; DB 1; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 0.082;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 ILLKLVAGF 9  
 Db 123 ILLKLVAGF 131  
  
 RESULT 3  
 YEAV\_ECOLI STANDARD; PRT; 481 AA.  
 AC P76252; P97208;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE PROBABLE TRANSPORT PROTEIN YEAV.  
 GN YEAV  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=97251358; PubMed=9097040;  
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
 RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,  
 RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,

RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,  
 RA Yamamoto Y., Horiuchi T.;  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 40.1-50.0 min region on the linkage map.";  
 RL DNA Res. 3:379-392(1996).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE BCCT FAMILY OF TRANSPORTERS (TC 2.33).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AE000274; AAC74871.1;  
 CC EMBL; D90823; BAA15596.1; ALT\_INIT.  
 CC EMBL; D90824; BAA15605.1; ALT\_INIT.  
 CC EcoGene; EGI3508; yeav.  
 CC InterPro; IPR000060;  
 CC Pfam; PF02028; BCCT; 1.  
 CC PROSITE; PS01303; BCCT; 1.  
 KW Hypothetical protein; transport; Transmembrane; Inner membrane.  
 FT TRANSMEM 32 52 POTENTIAL.  
 FT TRANSMEM 82 102 POTENTIAL.  
 FT TRANSMEM 137 157 POTENTIAL.  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT TRANSMEM 204 224 POTENTIAL.  
 FT TRANSMEM 258 278 POTENTIAL.  
 FT TRANSMEM 289 309 POTENTIAL.  
 FT TRANSMEM 348 368 POTENTIAL.  
 FT TRANSMEM 392 412 POTENTIAL.  
 FT TRANSMEM 418 438 POTENTIAL.  
 SQ SEQUENCE 481 AA; 52881 MW; 4F0DC3EB613F737A CRC64;  
  
 Query Match 82.9%; Score 34; DB 1; Length 481;  
 Best Local Similarity 77.8%; Pred. No. 8.7;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 ILLKLVAGF 9  
 Db 429 ILLKLVAGF 437  
  
 RESULT 4  
 PEVP\_LACDL STANDARD; PRT; 470 AA.  
 AC P45494;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE XAA-HIS DIPEPTIDASE (EC 3.4.13.3) (X-HIS DIPEPTIDASE) (AMINOACYL-  
 DE HISTIDINE DIPEPTIDASE) (CARNOSINASE).  
 GN PEVP.  
 OS Lactobacillus delbrueckii (subsp. lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=29397;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=DSM 7290 / WS87;  
 RX MEDLINE=95093606; PubMed=7528082;  
 RA Vongerichten K., Klein J., Matern H., Plapp R.;  
 RT "Cloning and nucleotide sequence analysis of pepV, a carnosinase gene  
 RT from Lactobacillus delbrueckii subsp. lactis DSM 7290, and partial  
 RT characterization of the enzyme.";  
 RL Microbiology 140:2591-2600(1994).  
 CC -1- FUNCTION: HAS ACTIVITY AGAINST BETA-ALANYL-DIPEPTIDES INCLUDING  
 CC CARNOSINE (BETA-ALANYL-HISTIDINE).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF XAA-|-HIS DIPEPTIDES.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A; ALSO KNOWN AS THE  
 CC ARGE/DAPE/ACY1/CPG2/YSCS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Z31377; CAA83252.1; -  
 CC MEROPS: M20.004; -  
 CC InterPro: IPR001261; -  
 CC InterPro: IPR002933; -  
 CC Pfam: PF01546; Peptidase\_M20; 1.  
 CC PROSITE: PS00758; ARGE\_DAPE\_CPG2\_1; 1.  
 CC PROSITE: PS00759; ARGE\_DAPE\_CPG2\_2; 1.  
 CC Hydrolase: Dipeptidase; Metalloprotease.  
 CC KW  
 CC SEQUENCE 470 AA; 51990 MW; 488117B4F33E4AB0 CRC64;  
 CC  
 CC Query Match 78.0%; Score 32; DB 1; Length 470;  
 CC Best Local Similarity 77.8%; Pred. No. 23;  
 CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC Qy 1 ILLKLVAGF 9  
 CC :||| |||  
 CC Db 131 MLLKKEAGF 139  
 CC  
 CC RESULT 5  
 CC MA6R YEAST  
 CC ID MA6R.YEAST STANDARD; PRT; 473 AA.  
 CC AC P10508;  
 CC DT 01-JUL-1989 (Rel. 11, Created)  
 CC DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 CC DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 CC DE MALTOSE FERMENTATION REGULATORY PROTEIN MAL6R.  
 CC GN MAL6R OR MAL63.  
 CC OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC OX NCBI\_TaxID=4932;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN-CARLSBERGENSIS / JM1901;  
 CC RX MEDLINE-89127146; PubMed-2851710;  
 CC RA Sollitt P., Marmur J.;  
 CC RT "Primary structure of the regulatory gene from the MAL6 locus of  
 CC Saccharomyces carlsbergensis.";  
 CC RL Mol. Gen. Genet. 213:56-62(1988).  
 CC [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE-89106267; PubMed-3145816;  
 CC RA Kim J., Michels C.A.;  
 CC RT "The MAL63 gene of Saccharomyces encodes a cysteine-zinc finger  
 CC protein.";  
 CC RL Curr. Genet. 14:319-323(1988).  
 CC  
 CC -1- FUNCTION: REGULATES THE COORDINATE TRANSCRIPTION OF STRUCTURAL  
 CC MAL6S (MALTAASE) AND MAL6T (MALTOSE PERMEASE) GENES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
 CC CLUSTER DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

CC EMBL: X12576; CAA31088.1; -  
 CC EMBL: M36537; AAA34755.1; -  
 CC PIR: S03814; RGBYM3.  
 CC TRANSFAC: T00480; -  
 CC SGD: L0001023; MAL63.  
 CC InterPro: IPR001138; -  
 CC Pfam: PF00172; Zn\_c1us; 1.  
 CC PRINTS: PR00054; FUNGALZNCYS.  
 CC PROSITE: PS00463; ZN2\_CYS6\_FUNGAL\_1; 1.  
 CC PROSITE: PS00463; ZN2\_CYS6\_FUNGAL\_2; 1.  
 CC Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 CC KW zinc; Metal-binding; Maltose metabolism; Multigene family.  
 CC FT DNA\_BIND 8 34 ZN(2)-CYS(6), FUNGAL-TYPE.  
 CC FT DOMAIN 41 49 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT CONFLICT 90 91 MISSING (IN REF. 2).  
 CC FT CONFLICT 206 206 MISSING (IN REF. 2).  
 CC SQ SEQUENCE 473 AA; 54895 MW; 679111C8871E643D CRC64;  
 CC  
 CC Query Match 78.0%; Score 32; DB 1; Length 473;  
 CC Best Local Similarity 62.5%; Pred. No. 23;  
 CC Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC Qy 2 LILKLVAGF 9  
 CC :||| |||  
 CC Db 204 LIKINGF 211  
 CC  
 CC RESULT 6  
 CC PORB.METJA  
 CC ID PORB.METJA STANDARD; PRT; 298 AA.  
 CC AC Q57714;  
 CC DT 30-MAY-2000 (Rel. 39, Created)  
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 CC DE PYRUVATE SYNTHASE SUBUNIT PORB (EC 1.2.7.1) (PYRUVATE OXIDOREDUCTASE  
 CC BETA CHAIN) (POR) (PYRUVIC-FERREDOXIN OXIDOREDUCTASE BETA SUBUNIT).  
 CC GN PORB OR MJ0266.  
 CC OS Methanococcus jannaschii.  
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 CC Methanococcus.  
 CC OX NCBI\_TaxID=2190;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 CC RX MEDLINE-96337999; PubMed-8688087;  
 CC RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 CC Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 CC Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 CC Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 CC Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 CC Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 CC Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 CC Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 CC "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 CC jannaschii.";  
 CC RL Science 273:1058-1073(1996).  
 CC  
 CC -1- CATALYTIC ACTIVITY: PYRUVATE + COA + OXIDIZED FERREDOXIN = ACETYL-  
 CC COA + CO(2) + REDUCED FERREDOXIN.  
 CC -1- SUBUNIT: HETEROTETRAMER OF ONE ALPHA, ONE BETA, ONE DELTA AND ONE  
 CC GAMMA CHAIN (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U67482; AAB98253.1; -  
 CC TIGR: MJ0266; -  
 CC

```

KW Oxidoreductase.
SQ SEQUENCE 298 AA; 33098 MW; D772C50EE8611A49 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 298;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8
DB 30 LLLKVAG 36
      |||||
      |||||

RESULT 7
SNA_A_BOVIN STANDARD; PRT; 295 AA.
AC P81125;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ALPHA-SOLUBLE NSF ATTACHMENT PROTEIN (SNAP-ALPHA).
GN NAPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN RATTUS
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=93205122; PubMed=8455721;
RA Whiteheart S.W., Griff I.C., Brunner M., Clary D.O., Mayer T.,
RA Burow S.A., Rothman J.E.;
RT "SNAP family of NSF attachment proteins includes a brain-specific
RT isoform.";
RL Nature 362:353-355(1993).
CC -!- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE ENDOPLASMIC
CC RETICULUM AND THE GOLGI APPARATUS.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN
CC BRAIN.
CC -!- SIMILARITY: TO YEAST SEC17.
CC InterPro; IPR000744; -.
DR Pfam; PF02071; NSF; 1.
DR PRINTS; PR00448; NSFATTACHMNT.
KW Transport; Protein transport; Endoplasmic reticulum; Golgi stack.
SQ SEQUENCE 295 AA; 33224 MW; 0453C5457D147E40 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 295;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLLKVAG 9
DB 165 LLLKVAG 171
      |||||
      |||||

RESULT 8
SNA_A_HUMAN STANDARD; PRT; 295 AA.
AC P54920;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALPHA-SOLUBLE NSF ATTACHMENT PROTEIN (SNAP-ALPHA).
GN NAPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN RATTUS
RP SEQUENCE FROM N.A.
RC TISSUE=Platelet;

KW Oxidoreductase.
SQ SEQUENCE 298 AA; 33098 MW; D772C50EE8611A49 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 298;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAG 8
DB 30 LLLKVAG 36
      |||||
      |||||

RESULT 7
SNA_A_BOVIN STANDARD; PRT; 295 AA.
AC P81125;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ALPHA-SOLUBLE NSF ATTACHMENT PROTEIN (SNAP-ALPHA).
GN NAPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN RATTUS
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=93205122; PubMed=8455721;
RA Whiteheart S.W., Griff I.C., Brunner M., Clary D.O., Mayer T.,
RA Burow S.A., Rothman J.E.;
RT "SNAP family of NSF attachment proteins includes a brain-specific
RT isoform.";
RL Nature 362:353-355(1993).
CC -!- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE ENDOPLASMIC
CC RETICULUM AND THE GOLGI APPARATUS.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN
CC BRAIN.
CC -!- SIMILARITY: TO YEAST SEC17.
CC InterPro; IPR000744; -.
DR Pfam; PF02071; NSF; 1.
DR PRINTS; PR00448; NSFATTACHMNT.
KW Transport; Protein transport; Endoplasmic reticulum; Golgi stack.
SQ SEQUENCE 295 AA; 33224 MW; 0453C5457D147E40 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 295;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLLKVAG 9
DB 165 LLLKVAG 171
      |||||
      |||||

RESULT 8
SNA_A_HUMAN STANDARD; PRT; 295 AA.
AC P54920;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALPHA-SOLUBLE NSF ATTACHMENT PROTEIN (SNAP-ALPHA).
GN NAPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN RATTUS
RP SEQUENCE FROM N.A.
RC TISSUE=Platelet;

MEDLINE=97413351; PubMed=9269766;
Lemons P.P., Chen D., Bernstein A.M., Bennett M.K., Whiteheart S.W.;
"Regulated secretion in platelets: identification of elements of the
platelet exocytosis machinery.";
Blood 90:1490-1500(1997).
(2)
REVIEWS TO 91; 133 AND 289.
Chen D., Shao H.P., Whiteheart S.W.;
Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE ENDOPLASMIC
RETICULUM AND THE GOLGI APPARATUS.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC PERIPHERAL MEMBRANE PROTEIN.
-!- SIMILARITY: TO YEAST SEC17.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U39412; AAC80170.1; -.
MIM; 603215; -.
InterPro; IPR000744; -.
Pfam; PF02071; NSF; 1.
PRINTS; PR00448; NSFATTACHMNT.
Transport; Protein transport; Endoplasmic reticulum; Golgi stack.
SEQUENCE 295 AA; 33246 MW; 1C691672A014F258 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 295;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLLKVAG 9
DB 165 LLLKVAG 171
      |||||
      |||||

RESULT 9
SNA_A_RAT STANDARD; PRT; 295 AA.
AC P54921;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ALPHA-SOLUBLE NSF ATTACHMENT PROTEIN (SNAP-ALPHA).
GN NAPA OR SNAP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN RATTUS
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=White adipose tissue;
RA Mitchell J.R.D., Mitchell M., Holman D., Oldfield S.;
Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE ENDOPLASMIC
RETICULUM AND THE GOLGI APPARATUS.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC PERIPHERAL MEMBRANE PROTEIN.
-!- SIMILARITY: TO YEAST SEC17.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X89968; CAA62005.1; -.
InterPro; IPR000744; -.

```



DR Pfam: PF02071; NSF; 1.  
 DR PRINTS: PR00448; NSPATTACHMNT.  
 KW Transport: Protein transport; Endoplasmic reticulum; Golgi stack.  
 SQ SEQUENCE 295 AA; 33192 MW; 662356859DF0BA3F CRC64;

Query Match 73.2%; Score 30; DB 1; Length 295;  
 Best Local Similarity 85.7%; Pred. No. 40;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 LKLVAGF 9  
 |||||:  
 Db 165 LKLVAGY 171

RESULT 10  
 YWCF\_BACSU STANDARD; PRT; 393 AA.  
 AC P39604; 1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE HYPOTHETICAL 43.3 KDA PROTEIN IN QOXD-VPR INTERGENIC REGION.  
 GN YWCF OR IPA-420.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95020537; PubMed=7934828;  
 RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,  
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,  
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,  
 RA Rapoport G., Danchin A.;  
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97  
 kb region from 325 degrees to 333 degrees.";  
 RL Mol. Microbiol. 10:371-384(1993).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE FTSW/RODA/SPOVE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X73124; CAAS1598.1; -;  
 DR EMBL: Z99123; CAB15838.1; -;  
 DR PIR: S39697; S39697.  
 DR Subtilist; BG10588; YWCF.  
 DR InterPro: IPR001182; -;  
 DR Pfam: PF01098; FTSW\_RODA\_SPOVE; 1.  
 DR PROSITE: PS00428; FTSW\_RODA\_SPOVE; 1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 17 37 POTENTIAL.  
 FT TRANSMEM 45 65 POTENTIAL.  
 FT TRANSMEM 75 95 POTENTIAL.  
 FT TRANSMEM 111 131 POTENTIAL.  
 FT TRANSMEM 146 166 POTENTIAL.  
 FT TRANSMEM 169 189 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT TRANSMEM 254 274 POTENTIAL.  
 FT TRANSMEM 285 305 POTENTIAL.  
 FT TRANSMEM 316 336 POTENTIAL.  
 FT TRANSMEM 342 362 POTENTIAL.  
 FT TRANSMEM 363 383 POTENTIAL.  
 SQ SEQUENCE 393 AA; 43274 MW; D84864CD368240FB CRC64;

Query Match 73.2%; Score 30; DB 1; Length 393;  
 Best Local Similarity 85.7%; Pred. No. 53;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 LLLKVAG 8  
 |||||:  
 Db 148 LLLKVIAG 154

RESULT 11  
 YXAM\_BACSU STANDARD; PRT; 399 AA.  
 AC P42112;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL 44.7 KDA PROTEIN IN ASNH-GNTR INTERGENIC REGION.  
 GN YXAM OR SI4MR.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE=96093926; PubMed=7584049;  
 RA Yoshida K.-I., Seki S., Fujimura M., Miwa Y., Fujita Y.;  
 RT "Cloning and sequencing of a 36-kb region of the Bacillus subtilis  
 genome between the gnt and lol operons.";  
 RL DNA Res. 2:61-69(1995).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=168 / BGSC1A1;  
 RA Fujita Y., Shibayama T., Ishio I., Aoyama D., Yoshida K.-I.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB005554; BAA21592.1; -;  
 DR EMBL: Z99124; CAB16029.1; -;  
 DR Subtilist; BG1115; YXAM.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 7 27 POTENTIAL.  
 FT TRANSMEM 33 53 POTENTIAL.  
 FT TRANSMEM 79 99 POTENTIAL.  
 FT TRANSMEM 131 151 POTENTIAL.  
 FT TRANSMEM 153 173 POTENTIAL.  
 FT TRANSMEM 208 228 POTENTIAL.  
 FT TRANSMEM 242 262 POTENTIAL.  
 FT TRANSMEM 296 316 POTENTIAL.  
 FT TRANSMEM 335 355 POTENTIAL.  
 FT TRANSMEM 357 377 POTENTIAL.  
 SQ SEQUENCE 399 AA; 44746 MW; 94AF85663222E72 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 399;  
 Best Local Similarity 55.6%; Pred. No. 54;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ILLLVAGF 9  
 :|||: :|||  
 Db 208 VLLIVVAGF 216

RESULT 12  
 OPPB\_MYCGE

```

ID  OPPB_MYCGE  STANDARD;  PRT;  407 AA.
AC  P47323;
DT  01-FEB-1996 (Rel. 33, Created)
DE  01-FEB-1996 (Rel. 33, Last sequence update)
DE  01-NOV-1997 (Rel. 35, Last annotation update)
DE  OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPB.
GN  OPPB OR MG077.
OS  Mycoplasma genitalium.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC  Mycoplasmataceae; Mycoplasma.
OX  NCBI_TaxID=2097;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 33530 / G-37;
RC  MEDLINE=96026346; PubMed=7569993;
RA  Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA  Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA  Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA  Nguyen D.T., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA  Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA  Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA  "The minimal gene complement of Mycoplasma genitalium.";
RL  Science 270:397-403(1995).
CC  -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC  FOR OLIGOPEPTIDES; PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF
CC  THE SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).
CC  -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC  -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC  PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPPBC
CC  SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U39688; AAC71295.1; -
DR  TIGR; MG077; -
DR  InterPro; IPR000515; -
DR  Pfam; PF00528; BPD_transp; 1.
DR  PROSITE; PS00402; BPD_TRANS_PNN_MEMBR; FALSE_NEG.
KW  Transport; Peptide transport; Transmembrane.
FT  TRANSMEM 9 29 POTENTIAL.
FT  TRANSMEM 101 121 POTENTIAL.
FT  TRANSMEM 134 154 POTENTIAL.
FT  TRANSMEM 179 199 POTENTIAL.
FT  TRANSMEM 239 259 POTENTIAL.
FT  TRANSMEM 288 308 POTENTIAL.
SQ  SEQUENCE 407 AA; 45464 MW; F25CBE4C5A4D4E80 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 407;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLKVAGF 9
DB 153 LLKLAGF 159
|||||

RESULT 13
YEDV_ECOLI
ID YEDV_ECOLI STANDARD; PRT; 452 AA.
AC P76339;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROBABLE SENSOR PROTEIN YEDV (EC 2.7.3.-).
GN YEDV.
OS Escherichia coli.

```

```

OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;
RC  MEDLINE=97426617; PubMed=9278503;
RA  Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RA  "The complete genome sequence of Escherichia coli K-12.";
RL  Science 277:1453-1474(1997).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12;
RC  MEDLINE=97251358; PubMed=9097040;
RA  Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA  Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA  Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA  Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA  Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
RA  Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RA  "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RA  corresponding to the 40.1-50.0 min region on the linkage map.";
RL  DNA Res. 3:379-392(1996).
CC  -!- FUNCTION: PROBABLE MEMBER OF A TWO-COMPONENT REGULATORY SYSTEM
CC  YEDW/YEDV. MAY ACTIVATE YEDW BY PHOSPHORYLATION.
CC  -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC  (PROBABLE).
CC  -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC  KINASES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE000288; AAC75034.1; -
DR  EMBL; D90835; BAA15795.1; -
DR  EcoGene; EGI4044; yedv.
DR  InterPro; IPR000410; -
DR  InterPro; IPR000658; -
DR  Pfam; PF00672; DUF5; 1.
DR  Pfam; PF00512; signal; 1.
KW  Hypothetical protein; Sensory transduction; Transference; Kinase;
KW  Transmembrane; Inner membrane; Phosphorylation.
FT  DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 10 30 POTENTIAL.
FT  DOMAIN 31 158 PERIPLASMIC (POTENTIAL).
FT  TRANSMEM 159 179 POTENTIAL.
FT  DOMAIN 180 452 CYTOPLASMIC (POTENTIAL).
FT  MOD_RES 245 245 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ  SEQUENCE 452 AA; 50849 MW; 172A6410C3CEBE8A CRC64;

Query Match 73.2%; Score 30; DB 1; Length 452;
Best Local Similarity 87.5%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKVAG 8
DB 15 ILLLSVAG 22
|||||

RESULT 14
YB96_MYCTU
ID YB96_MYCTU STANDARD; PRT; 334 AA.
AC P71777;
DT 15-DEC-1998 (Rel. 37, Created)

```

15-DEC-1998 (Rel. 37, Last sequence update)  
 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL 36.3 KDA PROTEIN RV1496.  
 GS RV1496 OR MTCY277.18.  
 ON Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -!- SIMILARITY: BELONGS TO THE ARKG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z79701; CAB02046.1; -  
 DR TuberculList; RV1496; -  
 KW Hypothetical protein; ATP-binding.  
 FT NP\_BIND 65 72 ATP (POTENTIAL)  
 SQ SEQUENCE 334 AA; 36256 MW; FFE57F8C7E6C38B5 CRC64;  
  
 Query Match 70.7%; Score 29; DB 1; Length 334;  
 Best Local Similarity 55.6%; Pred. No. 75;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 ILLKLVAGF 9  
 :|||: |||  
 Db 143 VVLEAAGF 151  
  
 RESULT 15  
 TIG\_NEIMA  
 ID TIG\_NEIMA STANDARD; PRT; 437 AA.  
 AC Q9JU32;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TRIGGER FACTOR (TF).  
 GN TIG OR NMA1526.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrell B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 RT meningitidis Z2491.";

RL Nature 404:502-506(2000).  
 CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT. ACTS AS A CHAPERONE BY  
 CC MAINTAINING THE NEWLY SYNTHESIZED PROTEIN IN AN OPEN CONFORMATION  
 CC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL162756; CAB84754.1; -  
 DR InterPro; IPR001179; -  
 DR PROSITE; PS00453; FKBP\_PPIASE\_1; FALSE\_NEG.  
 DR PROSITE; PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.  
 DR PROSITE; PS50059; FKBP\_PPIASE\_3; 1.  
 KW Cell division; Chaperone; Isomerase; Rotamase.  
 FT DOMAIN 163 248 PPIASE, FKBP-TYPE.  
 SQ SEQUENCE 437 AA; 48326 MW; F42BFF929751E616 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 4 LKVGAF 9  
 :|||||  
 Db 85 LKVGAF 90

Search completed: May 10, 2001, 10:15:22  
 Job time: 251 Sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:09:35 ; Search time 33.08 seconds  
(without alignments)  
18.697 Million cell updates/sec

Title: US-09-202-305-6

Perfect score: 41

Sequence: 1 ILLKLVAGF 9

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	264	2 F27579	T-cell receptor al
2	41	100.0	267	1 RWMSC8	T-cell receptor al
3	41	100.0	269	2 S57494	T-cell receptor al
4	41	100.0	269	2 S03715	T-cell receptor al
5	41	100.0	273	2 PL0063	T-cell receptor al
6	41	100.0	275	1 RWHUAC	T-cell receptor al
7	37	90.2	140	2 C44536	T-cell receptor al
8	37	90.2	271	2 A53268	T-cell receptor al
9	34	82.9	481	2 A64941	hypothetical prote
10	32	78.0	50	2 S05487	alpha-amylase (EC
11	32	78.0	139	2 A49054	T-cell receptor al
12	32	78.0	166	2 T33340	hypothetical prote
13	32	78.0	369	2 B83923	hypothetical prote
14	32	78.0	409	2 D75358	probable fosmidomy
15	32	78.0	470	2 S57902	peptidase V - Lact
16	32	78.0	473	1 RBYM3	regulatory protein
17	31	75.6	297	2 S75656	hypothetical prote
18	31	75.6	298	1 C64333	probable pyruvate
19	31	75.6	368	2 T36414	probable iron-side
20	31	75.6	375	2 F71234	probable Na+/H+-ex
21	31	75.6	392	2 F83279	probable MFS trans
22	31	75.6	418	2 A82737	probable transport
23	30	73.2	145	2 G72315	conserved hypothet
24	30	73.2	203	2 A74533	conserved hypothet
25	30	73.2	206	2 E71308	conserved hypothet
26	30	73.2	295	2 S58285	alpha-soluble NSF
27	30	73.2	295	2 G02238	alpha-SNAP - human
28	30	73.2	295	2 S32367	alpha-SNAP protein
29	30	73.2	369	2 T04917	hypothetical prote

30	30	73.2	375	2	C75201
31	30	73.2	389	2	B64574
32	30	73.2	393	2	S35697
33	30	73.2	399	2	D70072
34	30	73.2	407	2	E64208
35	30	73.2	408	2	H71813
36	30	73.2	408	2	B64708
37	30	73.2	452	2	D64961
38	30	73.2	470	2	S58826
39	30	73.2	478	2	C72658
40	30	73.2	572	2	F82595
41	30	73.2	787	2	T36582
42	30	73.2	1253	2	T14349
43	30	73.2	1253	2	T46248
44	29	70.7	44	2	E84151
45	29	70.7	58	2	S05489

#### ALIGNMENTS

##### RESULT 1

F27579

T-cell receptor alpha chain precursor V and C regions (TRA29) - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 08-Mar-1989 #sequence\_revision 30-Jun-1991 #text\_change 23-Jul-1999  
C:Accession: F27579

R.Morris, M.; Barclay, A.N.; Williams, A.F.

Immunogenetics 27, 174-179, 1988

A:Title: Analysis of T cell receptor beta chains in rat thymus, and rat C-alpha and C  
A:Reference number: A27578; MUID:88113841

A:Accession: F27579

A:Molecule type: mRNA

A:Residues: 1-264 <MOR>

A:Cross-references: EMBL:M18853; NID:q207163; PIDN:AAA42207.1; PID:q207164  
A:Note: The authors translated the codon CTC for residue 5 as Val

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: T-cell receptor

Query Match 100.0%; Score 41; DB 2; Length 264;  
Best Local Similarity 100.0%; Pred. No. 0.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9

Db 245 ILLKLVAGF 253

##### RESULT 2

RWMSC8

T-cell receptor alpha chain - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-May-1986 #sequence\_revision 01-Dec-2000 #text\_change 01-Dec-2000

C:Accession: S25117; A93345; A93344; B26492; A02132

R.Austrup, F.; Kodelja, V.; Kucharzik, T.; Kisch, E.

submitted to the EMBL Data Library, July 1992

A:Description: Characterization of idiotype-specific I-Ed-restricted T suppressor lym  
/c mice.

A:Reference number: S25117

A:Accession: S25117

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-267 <AUS>

A:Cross-references: EMBL:X67127

R.Saito, H.; Kranz, D.M.; Takagaki, Y.; Hayday, A.C.; Eisen, H.N.; Tonegawa, S.

Nature 312, 36-40, 1984

A:Title: A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes.

A:Reference number: A93345; MUID:85036635

A:Accession: A93345

A:Molecule type: mRNA

A:Residues: 130,'Y',132-267 <SAI>

A:Experimental source: strain BALB.B, clone pHD58, cytotoxic T lymphocyte

R.Chien, Y.; Becker, D.M.; Lindsten, T.; Okamura, M.; Cohen, D.I.; Davis, M.M.  
Nature 312, 31-35, 1984

A:Title: A third type of murine T-cell receptor gene.

A:Reference number: A93344; MUID:85036634

A:Accession: A93344

A:Molecule type: mRNA

A:Residues: 130, N', 132-267 <CHI>

A:Experimental source: hybridoma 2B4

R.Iimai, K.; Kanno, M.; Kimoto, H.; Shigemoto, K.; Yamamoto, S.; Taniguchi, M.  
Proc. Natl. Acad. Sci. U.S.A. 83, 8708-8712, 1986

A:Title: Sequence and expression of transcripts of the T-cell antigen receptor alpha-chain

A:Reference number: A94140; MUID:87041521

A:Accession: B26492

A:Molecule type: mRNA

A:Residues: 131-267 <IMA>

A:Experimental source: hybridoma 34S-281

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: glycoprotein; heterotrimer; receptor; T-cell; T-cell receptor; transmembrane

F:36-109/Domain: immunoglobulin homology <IMM>

F:130-241/Domain: C region #status predicted <CON>

F:242-262/Domain: transmembrane #status predicted <TM>

F:263-267/Domain: intracellular #status predicted <INT>

F:197,211,240/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 41; DB 1; Length 267;

Best Local Similarity 100.0%; Pred. No. 0.41;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILLKKVAGF 9

Db 248 ILLKKVAGF 256

RESULT 3

T-cell receptor alpha chain (V7S4AJ1759) - human

C:Species: Homo sapiens (man)

C>Date: 10-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 23-Jul-1999

C:Accession: S57494

R.Burrows, S.R.; Silins, S.L.; Moss, D.J.; Khanna, R.; Misko, I.S.; Arguet, V.P.

A:Description: T cell receptor repertoire for a viral epitope in humans is diversified b

A:Reference number: S57494

A:Accession: S57494

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-269 <BUR>

A:Cross-references: EMBL:249903; NID:q872124; PIDN:CAA90083.1; PID:g872125

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: T-cell receptor

Query Match 100.0%; Score 41; DB 2; Length 269;

Best Local Similarity 100.0%; Pred. No. 0.41;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILLKKVAGF 9

Db 250 ILLKKVAGF 258

RESULT 4

T-cell receptor alpha chain precursor (F5) - mouse

C:Species: Mus musculus (house mouse)

C>Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 23-Jul-1999

C:Accession: S03715

R.Palmer, M.S.; Bentley, A.; Gould, K.; Townsend, A.R.M.

A:Title: The T cell receptor from an Influenza-A specific murine CTL clone.

A:Reference number: S03715; MUID:89202046

A:Accession: S03715

A:Molecule type: mRNA

A:Residues: 1-269 <PAL>

C:Cross-references: EMBL:X14387; NID:g54509; PIDN:CAA32562.1; PID:g54510

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: T-cell receptor

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-269/Product: T-cell receptor alpha chain #status predicted <MAT>

Query Match 100.0%; Score 41; DB 2; Length 269;

Best Local Similarity 100.0%; Pred. No. 0.41;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILLKKVAGF 9

Db 250 ILLKKVAGF 258

RESULT 5

T-cell receptor alpha chain precursor V-J-C region - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Jul-1999

C:Accession: PL0063

R.Burns, F.R.; Li, X.; Shen, N.; Offner, H.; Chou, Y.K.; Vandenbark, A.A.; Heber-Katz

J. Exp. Med. 169, 27-39, 1989

A:Title: Both rat and mouse T cell receptors specific for the encephalitogenic determ

complex and encephalitogenic determinants being recognized are different.

A:Reference number: PL0063; MUID:89080488

A:Accession: PL0063

A:Molecule type: mRNA

A:Residues: 1-273 <BUR>

A:Cross-references: GB:X14318; NID:g57751; PIDN:CAA32498.1; PID:g762996

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: heterodimer; T-cell receptor

F:1-27/Domain: signal sequence (fragment) #status predicted <SIG>

F:28-273/Product: T-cell receptor alpha chain #status predicted <MAT>

F:28-118/Domain: V region #status predicted <VRE>

F:126-138/Domain: J region #status predicted <JRE>

F:139-273/Domain: C region #status predicted <CRE>

Query Match 100.0%; Score 41; DB 2; Length 273;

Best Local Similarity 100.0%; Pred. No. 0.42;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILLKKVAGF 9

Db 254 ILLKKVAGF 262

RESULT 6

RWHUAC

T-cell receptor alpha chain precursor - human

C:Species: Homo sapiens (man)

C>Date: 28-Feb-1986 #sequence\_revision 13-Mar-1997 #text\_change 22-Jun-1999

C:Accession: S18893; A02131

R.Hewitt, C.; Lamb, J.R.; Hill, M.; Owen, M.J.; Ohehir, R.O.; Hayball, J.

submitted to the EMBL Data Library, December 1991

A:Description: MHC independent clonal cell anergy by direct interaction of Staphyloco

A:Reference number: S18893

A:Accession: S18893

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-275 <HEW>

A:Cross-references: EMBL:X63455; NID:g36730; PIDN:CAA45055.1; PID:g36731

R.Yanagi, Y.; Chan, A.; Chin, B.; Minden, M.; Mak, T.W.

Proc. Natl. Acad. Sci. U.S.A. 82, 3430-3434, 1985

A:Title: Analysis of cDNA clones specific for human T cells and the alpha and beta ch

A:Reference number: A94046; MUID:85216514

A:Accession: A02131

A:Molecule type: mRNA

A:Residues: 134-275 <YAN>

A:Experimental source: clone pV14  
 C:Comment: This clone was isolated from a human leukemic T-cell line, Jurkat.  
 C:Comment: This alpha chain C region shows sequence homology to its beta chain counterpart  
 C:Comment: The gene corresponding to this mRNA is rearranged specifically in T-cells; it

C:Genetics:  
 A:Gene: GDB:TCRA  
 A:Cross-references: GDB:I20404; OMIM:186880  
 A:Map position: 14q11.2-14q11.2  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: heterotetramer; receptor; T-cell; transmembrane protein  
 F:134-250/Domain: C region <CON>  
 F:251-270/Domain: transmembrane #status predicted <TM>  
 F:271-275/Domain: intracellular #status predicted <INT>

Query Match 100.0%; Score 41; DB 1; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 ||||| ||

Db 256 ILLKLVAGF 264

## RESULT 7

C44536

T-cell receptor alpha chain C region - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 30-May-1997

C:Accession: C44536

R:Ishiguro, N.; Tanaka, A.; Shinagawa, M.

Immunogenetics 31, 57-60, 1990

A:Title: Sequence analysis of bovine T-cell receptor alpha chain.

A:Reference number: A45893; MUID:90129157

A:Accession: C44536

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-140 &lt;ISH&gt;

A:Cross-references: GB:D90030

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: T-cell receptor

Query Match 90.2%; Score 37; DB 2; Length 140;  
 Best Local Similarity 88.9%; Pred. No. 1.5;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 ||||| ||

Db 121 ILLKLVAGF 129

## RESULT 8

A53268

T-cell receptor alpha chain precursor - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 02-May-1994 #sequence\_revision 18-Nov-1994 #text\_change 23-Jul-1999

C:Accession: A53268

F:Hein, W.R.; Marcuz, A.; Fichtel, A.; Dudler, L.; Grossberger, D.

Immunogenetics 34, 39-41, 1991

A:Title: Primary structure of the sheep T-cell receptor alpha chain.

A:Reference number: A53268; MUID:91310085

A:Accession: A53268

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-271 &lt;HEI&gt;

A:Cross-references: GB:M55622; NID:gl66047; PIDN:AAA63518.1; PTD:gl66048

A:Note: sequence extracted from NCBI backbone (NCBI:44139, NCBIP:44140)

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: T-cell receptor

Query Match 90.2%; Score 37; DB 2; Length 271;  
 Best Local Similarity 88.9%; Pred. No. 2.9;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 ||||| ||

Db 252 ILLKLVAGF 260

## RESULT 9

A64941

hypothetical protein bl801 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 29-Sep-1999

C:Accession: A64941

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A:Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: A64941

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-481 &lt;BLAT&gt;

A:Cross-references: GB:AE000274; GB:U00096; NID:gl788089; PIDN:AAC74871.1; PID:gl7881

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: Escherichia coli probable carnitine transport protein

Query Match 82.9%; Score 34; DB 2; Length 481;  
 Best Local Similarity 77.8%; Pred. No. 22;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 ||||| ||

Db 429 ILLKLVAGF 437

## RESULT 10

S05487

alpha-amylase (EC 3.2.1.1) 2.53 precursor - wheat (fragment)

C:Species: Triticum aestivum (common wheat)

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 22-Jun-1999

C:Accession: S05487

R:Huttl, A.K.; Martienssen, R.A.; Baulcombe, D.C.

Mol. Gen. Genet. 214, 232-240, 1988

A:Title: Sequence heterogeneity and differential expression of the alpha-Amy-2 gene

A:Reference number: S05486; MUID:89181522

A:Accession: S05487

A&gt;Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-50 &lt;HUT&gt;

A:Cross-references: EMBL:X13577; NID:g21666; PIDN:CAA31914.1; PID:g21667

C:Genetics:

A:Gene: amy2

A:Map position: 7A

A:Introns: 29/3

C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds  
 A:Pathway: glycogen/starch degradation

C:Superfamily: wheat alpha-amylase; alpha-amylase core homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 78.0%; Score 32; DB 2; Length 50;  
 Best Local Similarity 77.8%; Pred. No. 6.5;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 ||||| ||

Db 12 LLLLLLVAGF 20

## RESULT 11

A49054  
 T-cell receptor alpha chain C region - pig (fragment)  
 C:Species: Sus scrofa domestica (domestic pig)  
 C>Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 23-Jul-1999  
 C:Accession: A49054  
 R:Thome, A.; Saalmuller, A.; Pfaff, E.  
 Eur. J. Immunol. 23, 1005-1010, 1993  
 A:Title: Molecular cloning of porcine T cell receptor alpha, beta, gamma and delta chain  
 A:Reference number: A49054; MUID:93238851  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-139 <THO>  
 A:Cross-references: GB:L21158; NID:g309796; PIDN:AAA65025.1; PID:g309797  
 A:Note: sequence extracted from NCBI backbone (NCBIP:130285)  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: T-cell receptor

Query Match 78.0%; Score 32; DB 2; Length 139;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 I:||||: ||  
 DB 120 ILLKLVAGF 128

## RESULT 12

T38340  
 hypothetical protein SPAC24B11.14 - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 04-Mar-2000  
 C:Accession: T38340  
 R:Odell, C.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: Z21786  
 A:Accession: T38340  
 A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA  
 A:Residues: 1-166 <ODE>  
 A:Cross-references: EMBL:Z67757; PIDN:CAB54976.1; GSPDB:GN00066; SPDB:SPAC24B11.14  
 A:Experimental source: strain 972h-; cosmid c24B11  
 C:Genetics:  
 A:Gene: SPDB:SPAC24B11.14  
 A:Map position: 1  
 A:Introns: 46/2; 65/2  
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC24B11.14.

Query Match 78.0%; Score 32; DB 2; Length 166;  
 Best Local Similarity 75.0%; Pred. No. 21;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLKVAGF 9  
 I:||||||  
 DB 65 LLVAVAGF 72

## RESULT 13

B83923  
 hypothetical protein B83923 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
 C:Accession: B83923  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20263314  
 A:Accession: B83923  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-369 <STO>  
 A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05905.1; GSPDB:G  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: B82186

Query Match 78.0%; Score 32; DB 2; Length 369;  
 Best Local Similarity 77.8%; Pred. No. 45;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 I:||||  
 DB 8 ILLILVAGF 16

## RESULT 14

D75358  
 probable fosmidomycin resistance protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: D75358  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
 , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: D75358  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-409 <WHI>  
 A:Cross-references: GB:AE002017; GB:AE000513; NID:g6459527; PIDN:AAF11300.1; PID:g645

A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRI1743  
 A:Map position: 1  
 C:Superfamily: fosmidomycin resistance protein

Query Match 78.0%; Score 32; DB 2; Length 409;  
 Best Local Similarity 77.8%; Pred. No. 50;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 I:||||  
 DB 127 ILLVAVAGF 135

## RESULT 15

S57902  
 peptidase V - Lactobacillus delbrueckii  
 C:Species: Lactobacillus delbrueckii  
 C>Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 24-Sep-1999  
 C:Accession: S57902  
 R:Vongerichten, K.F.; Klein, J.R.; Matern, H.; Plapp, R.  
 Microbiology 140, 2591-2600, 1994  
 A:Title: Cloning and nucleotide sequence analysis of pepV, a carnosinase gene from La  
 A:Reference number: S57902; MUID:95093606  
 A:Accession: S57902  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-470 <VON>  
 A:Cross-references: EMBL:Z31377; NID:g577568; PIDN:CAA83252.1; PID:g577569  
 C:Superfamily: peptidase V

Query Match 78.0%; Score 32; DB 2; Length 470;  
 Best Local Similarity 77.8%; Pred. No. 57;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
 I:|||| |||



Db 131 MLLKRGF 139

Search completed: May 10, 2001, 10:11:43  
Job time: 128 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:09:36 ; Search time 11.39 Seconds  
(without alignments)  
5.394 Million cell updates/sec

Title: US-09-202-305-6  
Perfect score: 41  
Sequence: 1 ILLKLVAGF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 39611 seqs, 6826361 residues

Total number of hits satisfying chosen parameters: 39611

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US05\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	277	5	US-09-822-827-906
2	33	80.5	164	5	US-09-739-449-11973
3	32	78.0	770	5	US-09-739-449-11556
4	31	75.6	510	6	US-60-248-505-934
5	30	73.2	99	1	PCT-US01-01339-3284
6	30	73.2	776	5	US-09-739-449-12367
7	29	70.7	403	5	US-09-739-449-8181
8	29	70.7	460	5	US-09-739-449-11412
9	28	68.3	129	5	US-09-828-769-282
10	28	68.3	216	5	US-09-828-644-74
11	28	68.3	333	5	US-09-739-449-8544
12	28	68.3	682	6	US-60-248-505-716
13	27	65.9	160	6	US-60-248-505-1006
14	27	65.9	190	5	US-09-739-449-8708
15	27	65.9	240	5	US-09-739-449-8640
16	27	65.9	318	5	US-09-739-449-9719
17	27	65.9	361	1	PCT-US01-04098A-1462
18	27	65.9	377	1	PCT-US01-01310-86
19	27	65.9	377	1	PCT-US01-01332-931
20	27	65.9	380	1	PCT-US01-12005-4
21	27	65.9	382	1	PCT-US01-12005-2
22	27	65.9	393	1	PCT-US01-04098A-3430
23	27	65.9	395	1	PCT-US00-35017A-1122
24	27	65.9	546	5	US-09-739-449-12091
25	27	65.9	636	5	US-09-423-844-175
26	26	63.4	42	5	US-09-809-391-591
27	26	63.4	43	5	US-09-809-391-446

28	26	63.4	125	5	US-09-739-449-12883	Sequence 12883, A
29	26	63.4	148	4	US-08-956-171C-5210	Sequence 5210, Ap
30	26	63.4	167	5	US-09-739-449-11387	Sequence 11387, A
31	26	63.4	171	6	US-60-248-505-877	Sequence 877, App
32	26	63.4	216	5	US-09-739-449-10190	Sequence 10190, A
33	26	63.4	224	1	PCT-US01-11988-1752	Sequence 1752, Ap
34	26	63.4	224	5	US-09-833-245-1752	Sequence 1752, Ap
35	26	63.4	270	5	US-09-739-449-10230	Sequence 10230, A
36	26	63.4	385	1	PCT-US01-11988-1754	Sequence 1754, Ap
37	26	63.4	385	5	US-09-833-245-1754	Sequence 1754, Ap
38	26	63.4	391	5	US-09-826-509-463	Sequence 463, App
39	26	63.4	394	5	US-09-739-449-10749	Sequence 10749, A
40	26	63.4	424	1	PCT-US01-11988-1753	Sequence 1753, Ap
41	26	63.4	424	5	US-09-833-245-1753	Sequence 1753, Ap
42	26	63.4	490	5	US-09-739-449-13176	Sequence 13176, A
43	26	63.4	566	5	US-09-739-449-10066	Sequence 10066, A
44	26	63.4	640	5	US-09-739-449-8451	Sequence 8451, Ap
45	26	63.4	1223	6	US-60-248-505-707	Sequence 707, App

## ALIGNMENTS

RESULT 1  
US-09-822-827-906  
; Sequence 906, Application US/09822827  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.534C1  
; CURRENT APPLICATION NUMBER: US/09/822.827  
; CURRENT FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 906  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-822-827-906

Query Match 100.0%; Score 41; DB 5; Length 277;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
DB 258 ILLKLVAGF 266

RESULT 2  
US-09-739-449-11973  
; Sequence 11973, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 11973  
; LENGTH: 164  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-11973

Query Match 80.5%; Score 33; DB 5; Length 164;  
Best Local Similarity 77.8%; Pred. No. 3;

```

Matches      7;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;

QY      1  ILLKLVAGF 9
      :||| ||||
Db      100 MLLNVAAGF 108

RESULT      3
US-09-739-449-11556
; Sequence 11556, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 11556
; LENGTH: 770
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-11556

Query Match      78.0%; Score 32; DB 5; Length 770;
Best Local Similarity 77.8%; Pred. No. 23;
Matches      7;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;

QY      1  ILLKLVAGF 9
      :||| ||||
Db      606 ILLRKLAF 614

RESULT      4
US-60-248-505-934
; Sequence 934, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 934
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Human
US-60-248-505-934

Query Match      75.6%; Score 31; DB 6; Length 510;
Best Local Similarity 44.4%; Pred. No. 24;
Matches      4;  Conservative      5;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1  ILLKLVAGF 9
      :||| ||||
Db      411 VLMLKISGY 419

RESULT      5
PCT-US01-01339-3284
; Sequence 3284, Application PC/TUS0101339
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01339

```

```

; CURRENT FILING DATE: 2001-03-17
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3284
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (40)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01339-3284

Query Match      73.2%; Score 30; DB 1; Length 99;
Best Local Similarity 85.7%; Pred. No. 7.4;
Matches      6;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

QY      3  LLKVAGF 9
      :||| ||||
Db      87 LLKVSAGF 93

RESULT      6
US-09-739-449-12367
; Sequence 12367, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 12367
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-12367

Query Match      73.2%; Score 30; DB 5; Length 776;
Best Local Similarity 77.8%; Pred. No. 59;
Matches      7;  Conservative      0;  Mismatches      2;  Indels      0;  Gaps      0;

QY      1  ILLKLVAGF 9
      :||| ||||
Db      648 ILLTAVAGF 656

RESULT      7
US-09-739-449-8181
; Sequence 8181, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 8181
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-8181

```

Query Match 70.7%; Score 29; DB 5; Length 403;  
Best Local Similarity 55.6%; Pred. No. 49;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
: : : : :  
Db 62 VLLSLAGF 70

RESULT 8  
US-09-739-449-11412  
; Sequence 11412, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(13490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 11412  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-11412

Query Match 70.7%; Score 29; DB 5; Length 460;  
Best Local Similarity 55.6%; Pred. No. 56;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
: : : : :  
Db 106 VLLTKLAGY 114

RESULT 9  
US-09-828-769-282  
; Sequence 282, Application US/09828769  
; GENERAL INFORMATION:  
; APPLICANT: Gearling, David P.  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
; FILE REFERENCE: Human Prostate Stromal Cell Library  
; CURRENT APPLICATION NUMBER: US/09/828,769  
; PRIOR FILING DATE: 2001-04-09  
; PRIOR APPLICATION NUMBER: 09/330,781  
; PRIOR FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 490  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 282  
; LENGTH: 129  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(29)  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(129)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-828-769-282

Query Match 68.3%; Score 28; DB 5; Length 129;  
Best Local Similarity 55.6%; Pred. No. 25;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
: : : : :  
Db 20 VLLAVVGF 28

RESULT 10  
US-09-828-644-74  
; Sequence 74, Application US/09828644  
; GENERAL INFORMATION:  
; APPLICANT: Vogeli, Gabriel  
; TITLE OF INVENTION: Novel G Protein-Coupled Receptors  
; FILE REFERENCE: 00196US1  
; CURRENT APPLICATION NUMBER: US/09/828,644  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,150  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/195,099  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/195,151  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/195,148  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/195,093  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/195,098  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/230,149  
; PRIOR FILING DATE: 2000-09-05  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 74  
; LENGTH: 216  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-828-644-74

Query Match 68.3%; Score 28; DB 5; Length 216;  
Best Local Similarity 66.7%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
: : : : :  
Db 1 ILLKLVAF 9

RESULT 11  
US-09-739-449-8544  
; Sequence 8544, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 09/514,000  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 8544  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-8544

Query Match 68.3%; Score 28; DB 5; Length 333;  
Best Local Similarity 44.4%; Pred. No. 64;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKLVAGF 9  
: : : : :  
Db 15 IIIIAGF 23

```
RESULT 12
US-60-248-505-716
; Sequence 716, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 716
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Human
US-60-248-505-716

Query Match      68.3%; Score 28; DB 6; Length 682;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LLLKVAGF 9
||| |||
Db 579 LLLSVGVF 586

RESULT 13
US-60-248-505-1006
; Sequence 1006, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1006
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Human
US-60-248-505-1006

Query Match      65.9%; Score 27; DB 6; Length 161;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKVAG 8
:|:|:|
Db 150 VLLMKVIG 157

RESULT 14
US-09-739-449-8708
; Sequence 8708, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351

; SEQ ID NO 8708
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-8708

Query Match      65.9%; Score 27; DB 5; Length 190;
Best Local Similarity 62.5%; Pred. No. 59;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ILLKVAG 8
:|:|:|
Db 115 VALIKVAG 122

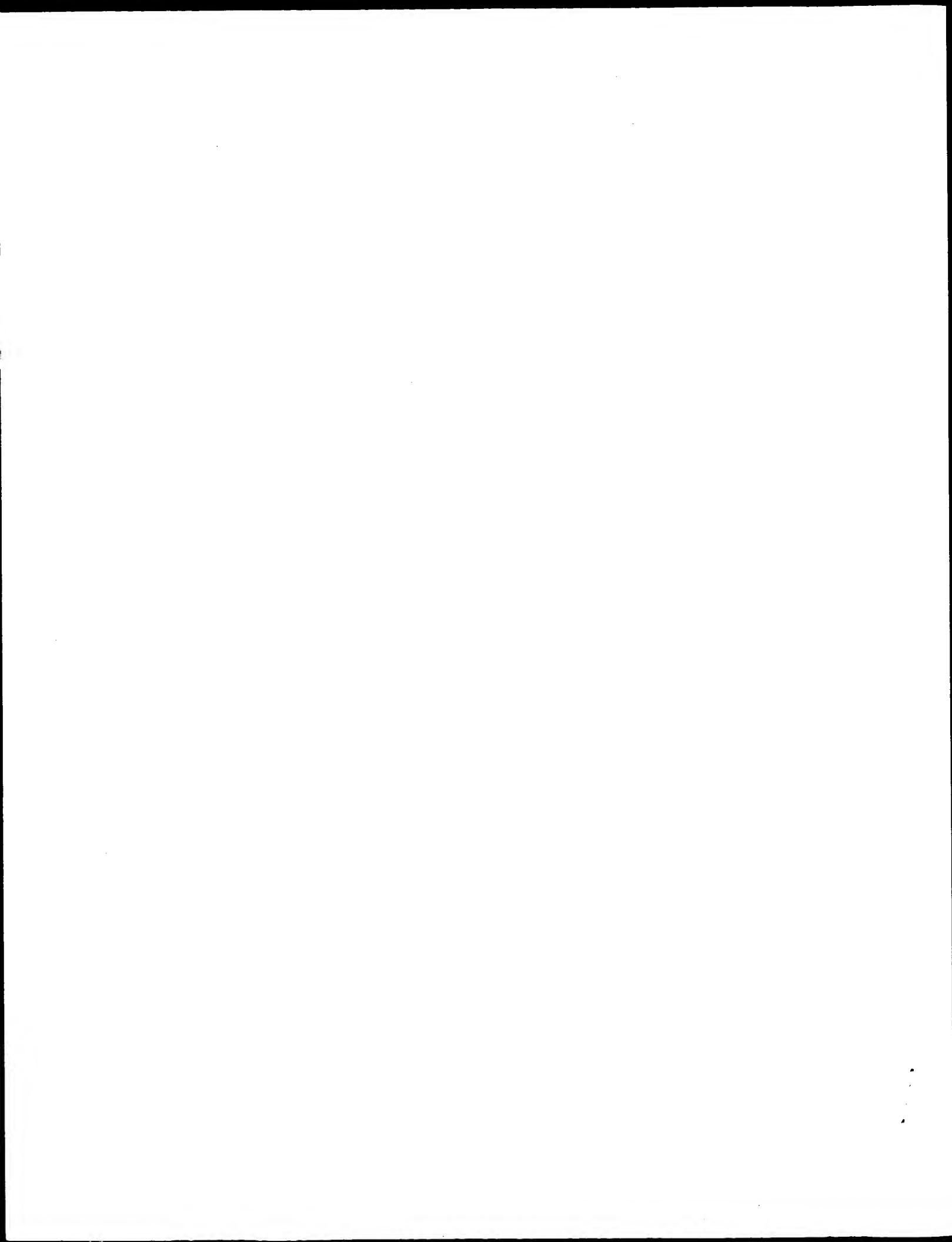
RESULT 15
US-09-739-449-8640
; Sequence 8640, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 8640
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-8640

Query Match      65.9%; Score 27; DB 5; Length 240;
Best Local Similarity 75.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LLLKVAGF 9
||| |||
Db 97 LLQKAAGF 104

Search completed: May 10, 2001, 10:14:00
Job time: 264 sec
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:11:44 ; Search time 33.08 seconds  
(without alignments)  
20.775 Million cell updates/sec

Title: US-09-202-305-13

Perfect score: 44

Sequence: 1 IIVTDVIATL 10

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 58722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	171	1 RWHUD1	T-cell surface gly
2	39	88.6	173	2 SI0375	T-cell surface gly
3	37	84.1	167	1 RWSHD3	T-cell surface gly
4	34	77.3	175	2 A39171	T-cell surface gly
5	34	77.3	235	2 JC2322	hypothetical 26.8k
6	33	75.0	187	2 S49755	adenine phosphorib
7	33	75.0	398	2 S40752	hypothetical prote
8	33	75.0	613	1 OQBE33	hypothetical prote
9	33	75.0	1010	1 PXZEP2	BRRF1 protein - hu
10	32	72.7	267	2 D81669	H+-transporting AT
11	32	72.7	309	2 C82186	tRNA pseudouridine
12	32	72.7	319	2 T22468	transcription regu
13	32	72.7	418	2 S33124	hypothetical prote
14	32	72.7	423	2 S73020	hypothetical prote
15	31	70.5	66	2 T44884	hypothetical prote
16	31	70.5	177	2 JS0097	hypothetical prote
17	31	70.5	180	2 E64230	hypothetical prote
18	31	70.5	227	2 T33094	hypothetical prote
19	31	70.5	275	1 JC1113	adenine phosphorib
20	31	70.5	321	2 T21156	probable orotate p
21	31	70.5	332	2 A65157	interleukin-2 rece
22	31	70.5	332	2 C64165	hypothetical prote
23	31	70.5	342	2 T28772	hypothetical prote
24	31	70.5	350	2 T21092	hypothetical prote
25	31	70.5	378	2 D82158	hypothetical prote
26	31	70.5	393	2 T13426	N-acetylglucosamin
27	31	70.5	547	2 A47178	hypothetical prote
28	31	70.5	669	2 JC5662	methyl-accepting t
29	31	70.5	799	2 T00052	hepatoma-derived g
					hypothetical prote

H+-transporting AT  
hypothetical prote  
hypothetical prote  
calgranulin c - pi  
adenine phosphorib  
probable adenine p  
adenine phosphorib  
hypothetical prote  
hypothetical prote  
uracil phosphoribo  
hypothetical prote  
probable apt - myc  
hypothetical prote  
hypothetical prote  
carbamate kinase (  
hypothetical prote

## ALIGNMENTS

### RESULT 1

RWHUD1

T-cell surface glycoprotein CD3 delta chain precursor - human

N:Alternate names: T-cell surface antigen T3/Leu 4 delta chain

C:Species: Homo sapiens (man)

C>Date: 04-Dec-1986 #sequence.revision 04-Dec-1986 #text.change 22-Jun-1999

C:Accession: A94706; A91040; B91040; A93347; S03952; A02443; A25109; A33003

R:van den Elsen, P.; Georgopoulos, A.; Shepley, B.A.; Orkin, S.; Terhorst, C.

Proc. Natl. Acad. Sci. U.S.A. 83, 2944-2948, 1986

A:Title: Exon/intron organization of the genes coding for the delta chains of the hum

A:Reference number: A94706; MUID:86206070

A:Accession: A94706

A:Molecule type: DNA

A:Residues: 1-171 <VAN>

A:Cross-references: GR:M12727; NID:gl79138; PIDN:AAA51792.1; PTD:gl79140

R:Tunnaciliffe, A.; Sims, J.E.; Rabbitts, T.H.

EMBO J. 5, 1245-1252, 1986

A:Title: T3-delta pre-mRNA is transcribed from a non-TATA promoter and is alternative

A:Reference number: A91040; MUID:86274627

A:Accession: A91040

A:Molecule type: DNA

A:Residues: 1-171 <TUN>

A:Cross-references: GB:X03934; NID:g37037; PIDN:CAA27573.1; PID:g755754

A:Accession: B91040

A:Molecule type: mRNA

A:Residues: 59-91, 'T', '93-171 <TU2>

A>Note: alternative splicing produces an mRNA missing the exon that codes for the tra

R:van den Elsen, P.; Shepley, B.A.; Borst, J.; Colligan, J.E.; Markham, A.F.; Orkin, S

Nature 312, 413-418, 1984

A:Title: Isolation of cDNA clones encoding the 20K T3 glycoprotein of human T-cell re

A:Reference number: A93347; MUID:85061572

A:Accession: A93347

A:Molecule type: mRNA

A:Residues: 1-171 <VA2>

A:Cross-references: GB:X01451; NID:g36774; PIDN:CAA55693.1; PID:g36775

R:Alexander, D.; Goris, J.; Marais, R.; Rothbard, J.; Merlevede, W.; Crumpton, M.J.

Eur. J. Biochem. 181, 55-65, 1989

A:Title: Dephosphorylation of the human T lymphocyte CD3 antigen.

A:Reference number: S03951; MUID:89231703

A:Accession: S03952

A>Status: preliminary

A:Molecule type: protein

A:Residues: 128-171 <ALE>

C:Comment: This protein is one of five that make up the T-cell receptor/T3 complex.

C:Genetics:

A:Gene: GDB:CD3D

A:Cross-references: GDB:120578; OMIM:186790

A:Map position: 11q23.3-11q23.3

A:Introns: 19/1; 92/1; 136/1; 150/3

C:Superfamily: T-cell surface glycoprotein CD3 delta chain; immunoglobulin homology

C:Keywords: glycoprotein; receptor; T-cell; transmembrane protein

F,1-21/Domain: signal sequence #status predicted <SIG>

F:22-171/Product: T-cell surface glycoprotein CD3 delta chain #status predicted <MAT>  
 F:22-103/Domain: extracellular #status predicted <EXT>  
 F:30-75/Domain: immunoglobulin homology <IMM>  
 F:104-125/Domain: transmembrane #status predicted <TM>  
 F:126-171/Domain: intracellular #status predicted <INT>  
 F:37-73/Disulfide bonds: #status predicted  
 F:38,74/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 44; DB 1; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 0.1;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIVTDVIATL 10  
 :|:|:|:|:|:|  
 Db 107 IIIVTDVIATL 116

## RESULT 2

T-cell surface glycoprotein CD3 delta chain precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 18-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
 C:Accession: S10975  
 R:Davies, J.D.; Mueller, D.; Wilson, D.B.; Gold, D.P.  
 Nucleic Acids Res. 18, 4617, 1990  
 A:Title: Nucleotide sequence of a cDNA encoding the rat T3 delta chain.  
 A:Reference number: S10975; MUID:90356424  
 A:Accession: S10975  
 A:Molecule type: mRNA  
 A:Residues: 1-173 <DAV>  
 A:Cross-references: EMBL:X53430; NID:955913; PIDN:CAA37521.1; PID:955914  
 C:Superfamily: T-cell surface glycoprotein CD3 delta chain; immunoglobulin homology  
 C:Keywords: glycoprotein; transmembrane protein  
 F:30-75/Domain: immunoglobulin homology <IMM>

Query Match 88.6%; Score 39; DB 2; Length 173;  
 Best Local Similarity 70.0%; Pred. No. 1.1;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIVTDVIATL 10  
 :|:|:|:|:|  
 Db 107 VIITDLIATL 116

## RESULT 3

T-cell surface glycoprotein CD3 delta chain precursor - sheep  
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 22-Jun-1999  
 C:Accession: B43547; S11725  
 R:Hein, W.R.; Tunncliffe, A.  
 Eur. J. Immunol. 20, 1505-1511, 1990  
 A:Title: Characterization of the CD3 gamma and delta invariant subunits of the sheep T cell receptor.  
 A:Reference number: A43547; MUID:90353383  
 A:Accession: B43547  
 A:Molecule type: mRNA  
 A:Residues: 1-167 <HEI>  
 A:Cross-references: EMBL:X52993; NID:91241; PIDN:CAA37182.1; PID:g1242  
 A:Note: submitted to the EMBL data library, May 1990  
 C:Superfamily: T-cell surface glycoprotein CD3 delta chain; immunoglobulin homology  
 C:Keywords: glycoprotein; receptor; T-cell; transmembrane protein  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-167/Product: T-cell surface glycoprotein CD3 delta chain #status predicted <MAT>  
 F:22-97/Domain: extracellular #status predicted <EXT>  
 F:98-122/Domain: transmembrane #status predicted <TM>  
 F:38,55/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.1%; Score 37; DB 1; Length 167;  
 Best Local Similarity 60.0%; Pred. No. 2.7;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIVTDVIATL 10  
 :|:|:|:|:|  
 Db 103 LIITDLIATV 112

## RESULT 4

T-cell surface glycoprotein CD3 delta and gamma chain homolog, precursor - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 07-Feb-1992 #sequence\_revision 07-Feb-1992 #text\_change 23-Jul-1999  
 C:Accession: A39171  
 R:Bernot, A.; Auffray, C.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2550-2554, 1991  
 A:Title: Primary structure and ontogeny of an avian CD3 transcript.  
 A:Reference number: A39171; MUID:91172844  
 A:Accession: A39171  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-175 <BER>  
 A:Cross-references: GB:M59925; NID:g211433; PIDN:AAA48660.1; PID:g211434  
 C:Superfamily: T-cell surface glycoprotein CD3 delta chain; immunoglobulin homology  
 C:Keywords: glycoprotein

Query Match 77.3%; Score 34; DB 2; Length 175;  
 Best Local Similarity 60.0%; Pred. No. 12;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIIVTDVIATL 10  
 :|:|:|:|:|  
 Db 112 IIVADVAVATV 121

## RESULT 5

hypothetical 26.8K protein - plasmid pAH4  
 C:Species: plasmid pAH4  
 A:Note: host Acetobacter BRR2001, cellulose producing  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 05-Nov-1999  
 C:Accession: JC2322  
 R:Tonouchi, N.; Tsuchida, T.; Yoshinaga, F.; Horinouchi, S.; Beppu, T.  
 Biosci. Biotechnol. Biochem. 58, 1899-1901, 1994  
 A:Title: A host-vector system for a cellulose-producing Acetobacter strain.  
 A:Reference number: JC2319; MUID:95072320  
 A:Accession: JC2322  
 A:Molecule type: DNA  
 A:Residues: 1-235 <HOS>  
 A:Cross-references: DDBJ:D30784; NID:g598439; PIDN:BA06447.1; PID:dl007018; PID:g999  
 C:Genetics:  
 A:Genome: plasmid

Query Match 77.3%; Score 34; DB 2; Length 235;  
 Best Local Similarity 66.7%; Pred. No. 16;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIIVTDVIAT 9  
 :|:|:|:|:|  
 Db 80 IVTDDVAVAT 88

## RESULT 6

adenine phosphoribosyltransferase (EC 2.4.2.7) 1 - yeast (Saccharomyces cerevisiae)  
 S49755  
 N:Alternate names: protein YML022w  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 13-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 18-Jun-1999  
 C:Accession: S49755; JC4213; S55715  
 R:Badcock, K.; Churcher, C.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S49741  
 A:Accession: S49755

A:Molecule type: DNA  
A:Residues: 1-187 <BAD>  
A:Cross-references: EMBL:Z46659; NID:q575680; PIDN:CAA86633.1; PID:q575697; MIPS:YML022w  
R:Alfonzo, J.D.; Sahota, A.; Deeley, M.C.; Ranjekar, P.; Taylor, M.W.  
Gene 161, 81-85, 1995  
A:Title: Cloning and characterization of the adenine phosphoribosyltransferase-encoding  
A:Reference number: JC4213; MUID:95369738  
A:Accession: JC4213  
A:Molecule type: DNA  
A:Residues: 1-36, 'L', 38-101, 'S', 103-187, 'SNNVSRDPYSSFLSAGFKQIFLTSPYSLNVV' <ALF>  
A:Cross-references: GB:U16781  
C:Genetics:  
A:Gene: SGD:APT1  
A:Cross-references: SGD:S0004484; MIPS:YML022w  
A:Map position: 13L  
C:Superfamily: adenine phosphoribosyltransferase  
C:Keywords: glycosyltransferase; pentosyltransferase; transmembrane protein  
F:72-88/Domain: transmembrane #status predicted <TM>  
F:125-139/Region: 5'-phosphoribosylpyrophosphate binding #status predicted

Query Match 75.0%; Score 33; DB 2; Length 187;  
Best Local Similarity 77.8%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIVTDVIAT 9  
||| |  
Db 126 IIVDDIIAT 134

RESULT 7  
S40752  
Hypothetical protein C15H7.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997  
C:Accession: S40752  
R:Smith, A.  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S40750  
A:Accession: S40752  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-398 <SMI>  
A:Cross-references: EMBL:Z22173; NID:q297944; PID:q297947  
C:Genetics:  
A:Introns: 123/1; 311/3

Query Match 75.0%; Score 33; DB 2; Length 398;  
Best Local Similarity 55.6%; Pred. No. 45;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 IIVTDVIATL 10  
||| |  
Db 295 VVTDLVATM 303

RESULT 8  
QBEB33  
BRRF1 protein - human herpesvirus 4 (strain B95-8)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C:Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 16-Jul-1999  
C:Accession: E43043; A03775; S33025  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-45, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
A:Reference number: A93065; MUID:85035713  
A:Accession: E43043  
A:Molecule type: DNA  
A:Residues: 1-613 <BAN>  
A:Cross-references: EMBL:Y01555; NID:q59074; PIDN:CAA24820.1; PID:gl334884  
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H  
Nature 310, 207-211, 1984

A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667  
A:Contents: annotation; protein coding region  
C:Superfamily: varicella-zoster virus gene 54 protein

Query Match 75.0%; Score 33; DB 1; Length 613;  
Best Local Similarity 77.8%; Pred. No. 71;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 IIVTDVIATL 10  
||| |  
Db 538 LVTDVYATL 546

RESULT 9  
PZP2P  
H-transporting ATPase (EC 3.6.1.35) 2, plasma membrane [validated] - fission yeast  
N:Alternate names: proton pump  
C:Species: Schizosaccharomyces pombe  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 01-Dec-2000  
C:Accession: A40945; T40840; T40949  
R:Chislain, M.; Goffeau, A.  
J. Biol. Chem. 266, 18276-18279, 1991  
A:Title: The pma1 and pma2 H(+)-ATPases from Schizosaccharomyces pombe are functional  
A:Reference number: A40945; MUID:92011563  
A:Accession: A40945  
A:Molecule type: DNA  
A:Residues: 1-1010 <GHI>  
A:Cross-references: GB:M60471; NID:gl73430; PIDN:AAA35325.1; PID:gl73431  
R:Oliver, K.; Harris, D.; Wood, V.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: 221950  
A:Accession: T40840  
A:Molecule type: DNA  
A:Residues: 1-573 <OLI>  
A:Cross-references: EMBL:AL023518; PIDN:CAA19899.1; GSPDB:GN00068; SPDB:SPCC1020.01c  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
submitted to the EMBL Data Library, February 1999  
A:Reference number: 221940  
A:Accession: T40949  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 466-1010 <WOO>  
A:Cross-references: EMBL:AL035592; PIDN:CAB38157.1; GSPDB:GN00068; SPDB:SPCC1393.01  
A:Experimental source: strain 972h(-)  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
submitted to the EMBL Data Library, February 1999  
A:Reference number: 221940  
A:Accession: T40949  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 466-1010 <WOO>  
A:Cross-references: EMBL:AL035592; PIDN:CAB38157.1; GSPDB:GN00068; SPDB:SPCC1393.01  
A:Experimental source: strain 972h(-); cosmid cl393  
C:Comment: This ATPase transports protons across the plasma membrane to regulate intr  
C:Comment: In addition to transient phosphorylation of the active site Asp residue, t  
membrane.  
C:Genetics:  
A:Gene: PMA2; SPDB:SPCC1393.01  
A:Map position: 3L  
A:Introns: #status absent  
C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding doma  
C:Keywords: ATP; hydrogen ion transport; hydrolase; phosphoprotein; transmembrane pro  
F:202-222/Domain: transmembrane #status predicted <TM1>  
F:227-246/Domain: transmembrane #status predicted <TM2>  
F:379-400/Domain: transmembrane #status predicted <TM3>  
F:411-428/Domain: transmembrane #status predicted <TM4>  
F:614-785/Domain: ATPase nucleotide-binding domain homology <ATN>  
F:807-825/Domain: transmembrane #status predicted <TM5>  
F:841-859/Domain: transmembrane #status predicted <TM6>  
F:913-933/Domain: transmembrane #status predicted <TM7>  
F:941-965/Domain: transmembrane #status predicted <TM8>  
F:464/Active site: Asp (aspartylphosphate intermediate) #status predicted

Query Match 75.0%; Score 33; DB 1; Length 1010;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10

Db 917 VLVVDIATL 926  
::| |::| |

## RESULT 10

D81669  
tRNA pseudouridine synthase A TC0748 [imported] - Chlamydia muridarum (strain Nigg)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 21-Jul-2000  
C:Accession: D81669  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: D81669  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-267 <TET>  
A:CROSS-references: GB:AE002343; GB:AE002160; NID:g7190772; PIDN:AAF39555.1; PID:g7190772  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:  
A:Gene: TC0748  
C:Superfamily: tRNA-pseudouridine synthase I

## Query Match

Best Local Similarity 72.7%; Score 32; DB 2; Length 267;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IIIVTDVIAT 9

Db 91 IVIRDVIAT 99

## RESULT 11

C82186  
transcription regulator LysR family VCL1561 [imported] - Vibrio cholerae (group O1 strain  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Sep-2000  
C:Accession: C82186  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.  
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: C82186

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-309 <HEI>

A:CROSS-references: GB:AE004233; GB:AE003852; NID:g9656055; PIDN:AAF94715.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCL1561

A:Map position: 1

## Query Match

Best Local Similarity 72.7%; Score 32; DB 2; Length 309;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IIIVTDVIATL 10

Db 240 LINTDLIATL 249

## RESULT 12

T22468

hypothetical protein F49H6.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000

C:Accession: T22468

R:Lloyd, C.  
submitted to the EMBL Data Library, November 1996

A:Reference number: Z19567

A:Accession: T22468

A:Status: preliminary; translated from GB/EMBL/DDRJ

A:Molecule type: DNA

A:Residues: 1-319 <WIL>

A:CROSS-references: EMBL:Z81545; PIDN:CAR04443.1; GSPDB:GN00023; CESP:F49H6.10

A:Experimental source: clone F49H6

C:Genetics:

A:Gene: CESP:F49H6.10

A:Map position: 5

A:Introns: 48/3; 93/2; 167/1

C:Superfamily: Caenorhabditis elegans hypothetical protein K02H11.4

## Query Match

Best Local Similarity 72.7%; Score 32; DB 2; Length 319;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IIIVTDVIAT 9

Db 259 IIITDVIAT 267

## RESULT 13

S31124

hypothetical protein F59B2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997

C:Accession: S31124

R:Sulston, J.; Du, Z.; Thomas, K.; Wilson, R.; Hillier, L.; Staden, R.; Halloran, N.;

awkins, T.; Ainscough, R.; Waterston, R.

submitted to the EMBL Data Library, November 1991

A:Description: The C. elegans sequencing project: A beginning.

A:Reference number: S31122

A:Accession: S31124

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-418 <SUL>

A:CROSS-references: EMBL:Z11505; NID:g6718; PID:g6721

C:Genetics:

A:Introns: 63/3; 82/1; 131/3; 150/1; 251/3; 374/1

## Query Match

Best Local Similarity 72.7%; Score 32; DB 2; Length 418;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IIIVTDVIATL 10

Db 302 VLVTDIAAAL 311

## RESULT 14

S73020

hypothetical protein L518\_F2.48 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 22-Oct-1999

C:Accession: S73020

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid L518.

A:Reference number: S72580

A:Accession: S73020

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-423 <SMI>

A:CROSS-references: EMBL:U00023; NID:g467194; PIDN:AAA17363.1; PID:g467206

C:Genetics:

A:Start codon: GTG

Query Match 72.7%; Score 32; DB 2; Length 423;  
Best Local Similarity 55.6%; Pred. No. 76;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 2 IVTDVIATL 10  
      : ||::|||  
Db 304 LATDIVATL 312

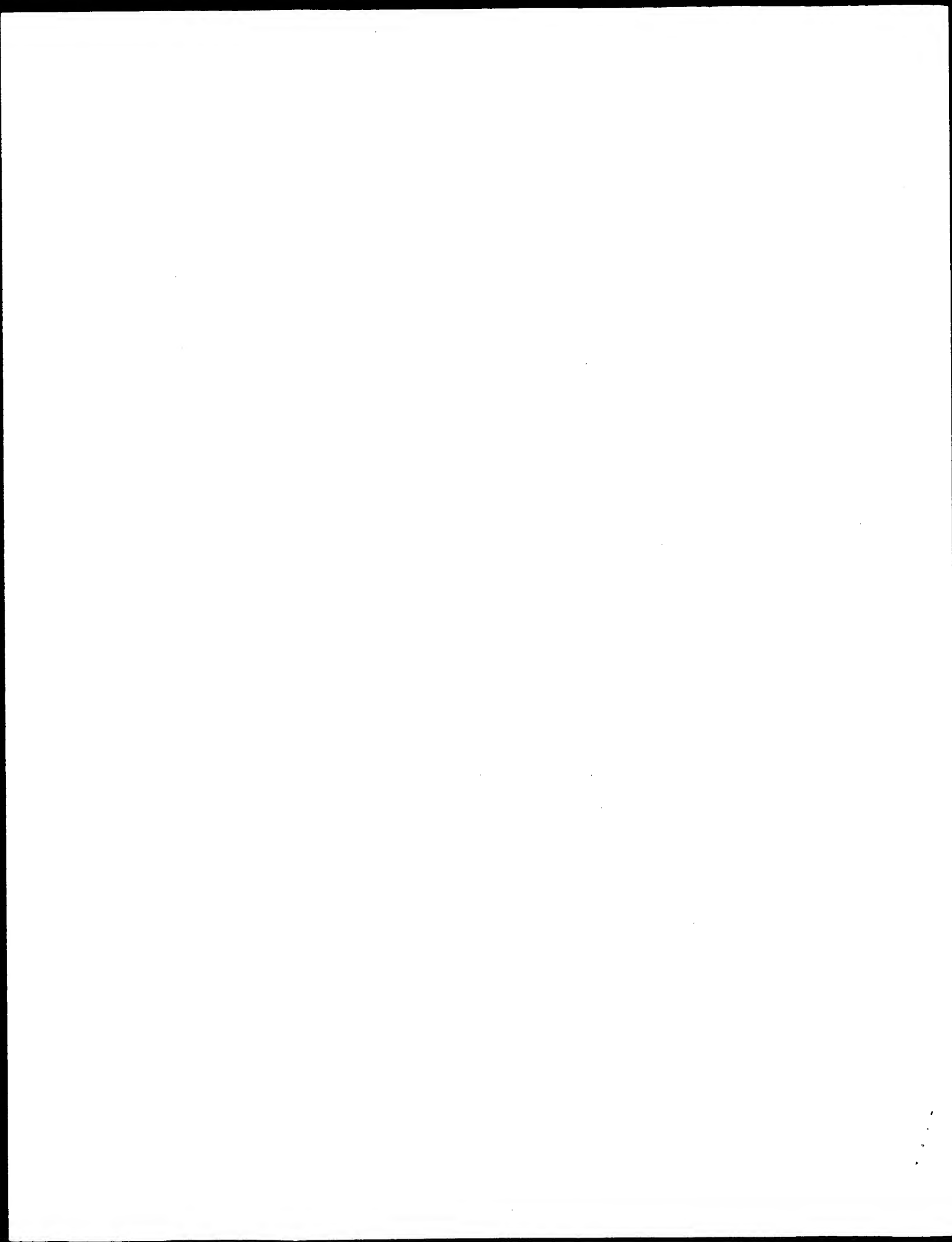
## RESULT 15

T44884  
hypothetical protein MLCB22.0lc [imported] - Mycobacterium leprae (fragment)  
C:Species: Mycobacterium leprae  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 02-Sep-2000  
C:Accession: T44884  
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: 222864  
A:Accession: T44884  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-66 <PAR>  
A:Cross-references: EMBL:Z98741; PIDN:CAB11366.1  
A:Experimental source: cosmid B22  
C:Genetics:  
A:Note: MLCB22.0lc  
C:Superfamily: conserved hypothetical protein HI0376

Query Match 70.5%; Score 31; DB 2; Length 66;  
Best Local Similarity 66.7%; Pred. No. 17;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 IIIVTDVIAT 9  
      : |||||  
Db 14 VILTDVAAT 22

Search completed: May 10, 2001, 10:11:45  
Job time: 130 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:15:23 ; Search time 20.13 seconds  
(without alignments)  
17.017 Million cell updates/sec

Title: US-09-202-305-13  
Perfect score: 44  
Sequence: 1 IIVTDVIATL 10

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	171	1	CD3D_HUMAN
2	39	88.6	173	1	CD3D_RAT
3	37	84.1	167	1	CD3D_SHEEP
4	33	75.0	187	1	APTL_YEAST
5	33	75.0	398	1	YK13_CAEEL
6	33	75.0	613	1	UL06_EBV
7	33	75.0	916	1	PMAL_AJECA
8	33	75.0	1010	1	PMAL_SCHPO
9	32	72.7	418	1	NAGA_CAEEL
10	32	72.7	899	1	PMAL_KLUULA
11	32	72.7	1158	1	ALAL_ARATH
12	31	70.5	176	1	TPF1_TREPA
13	31	70.5	176	1	TYF1_TREPE
14	31	70.5	180	1	APTL_MYCGE
15	31	70.5	275	1	IL2A_SHEEP
16	31	70.5	332	1	YIAK_ECOLI
17	31	70.5	332	1	YIAK_HAEIN
18	31	70.5	397	1	ORZA_DROME
19	31	70.5	547	1	MCPK_SALTY
20	31	70.5	920	1	PMAL_ZYGRO
21	30	68.2	91	1	S112_PIG
22	30	68.2	181	1	APT2_YEAST
23	30	68.2	183	1	APTL_ARATH
24	30	68.2	223	1	APTL_MYCTU
25	30	68.2	255	1	CAPC_STAUA
26	30	68.2	310	1	ARCC_HAEIN
27	30	68.2	342	1	YISS_BACSU
28	30	68.2	458	1	YCG9_YEAST
29	30	68.2	519	1	C6G2_DROME
30	30	68.2	582	1	YK85_YEAST
31	30	68.2	895	1	PMAL_CANAL
32	29	65.9	78	1	CBX2_HUMAN
33	29	65.9	172	1	APT_SVNY3

#### ALIGNMENTS

```

RESULT 1
CD3D_HUMAN
ID CD3D_HUMAN STANDARD; PRT; 171 AA.
AC P04234.
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE T-CELL SURFACE GLYCOPROTEIN CD3 DELTA CHAIN PRECURSOR (T-CELL RECEPTOR
DE T3 DELTA CHAIN).
GN CD3D OR T3D.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86206070; PubMed=2939461;
RA van den Elsen P., Georgopoulos K., Shepley B.-A., Orkin S.,
RA Terhorst C.;
RT "Exon/intron organization of the genes coding for the delta chains of
RT the human and murine T-cell receptor/T3 complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:2944-2948(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061572; PubMed=6095101;
RA van den Elsen P., Shepley B.-A., Borst J., Coligan J.E., Markham A.F.,
RA Orkin S., Terhorst C.;
RT "Isolation of cDNA clones encoding the 20K T3 glycoprotein of human
RT T-cell receptor complex.";
RL Nature 312:413-418(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=86274627; PubMed=3488209;
RA Tunnaciffe A., Sims J.E., Rabbitts T.H.;
RT "T3 delta pre-mRNA is transcribed from a non-TATA promoter and is
RT alternatively spliced in human T cells.";
RL EMBL J. 5:1245-1252(1986).
CC -!- FUNCTION: THE CD3 COMPLEX MEDIATES SIGNAL TRANSDUCTION.
CC -!- SUBUNIT: THE TCR/CD3 COMPLEX OF T LYMPHOCYTES CONSISTS OF EITHER
CC A TCR ALPHA/BETA OR TCR GAMMA/DELTA HETERODIMER COEXPRESSED AT THE
CC CELL SURFACE WITH THE INVARIANT SUBUNITS OF CD3 LABELED GAMMA,
CC DELTA, EPSILON, ZETA, AND ETA.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE GAMMA,
CC DELTA, AND EPSILON CHAINS ARE HIGHLY SIMILAR.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD3d entry.
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd3delta.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: X03934; CAA27573.1; -
DR EMBL: M12727; AAA51792.1; -
DR EMBL: M12726; AAA51792.1; JOINED.
DR EMBL: X01451; CAA25683.1; -
DR PIR: A02245; RWHUD1.
DR MIM: 186790; -
KW Immunoglobulin domain; T-cell; Receptor; Transmembrane; Glycoprotein;
KW Signal.
FT SIGNAL. 1 21
FT CHAIN 22 171
FT FT T-CELL SURFACE GLYCOPROTEIN CD3 DELTA
FT FT CHAIN.
FT FT
FT FT
FT DOMAIN 22 100
FT TRANSMEM 101 127
FT POTENTIAL.
FT DOMAIN 128 171
FT CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 38 38
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 74 74
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 74 74
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 171 AA; 18929 MW; 6C1F248150186D21 CRC64;

Query Match 100.0%; Score 44; DB 1; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.054;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10
| | | | | | | |
Db 107 IIVTDVIATL 116

RESULT 2
CD3D_RAT CD3D_RAT STANDARD; PRT; 173 AA.
AC P19377;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE T-CELL SURFACE GLYCOPROTEIN CD3 DELTA CHAIN PRECURSOR (T-CELL RECEPTOR
DE T3 DELTA CHAIN).
GN CD3D OR T3D.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WISTAR;
RX MEDLINE=90356424; PubMed=2143819;
RA Davies J.D., Mueller D., Wilson D.B., Gold D.P.;
RA "Nucleotide sequence of a cDNA encoding the rat T3 delta chain.";
RL Nucleic Acids Res. 18:4617-4617(1990).
CC -!- FUNCTION: THE CD3 COMPLEX MEDIATES SIGNAL TRANSDUCTION.
CC A TCR ALPHA/BETA OR TCR GAMMA/DELTA HETERODIMER COEXPRESSED AT THE
CC CELL SURFACE WITH THE INVARIANT SUBUNITS OF CD3 LABELED GAMMA,
CC DELTA, EPSILON, ZETA, AND ETA.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE GAMMA,
CC DELTA, AND EPSILON CHAINS ARE HIGHLY SIMILAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53430; CAA37521.1; -
CC PIR: S10975; S10975
CC Immunoglobulin domain; T-cell; Receptor; Transmembrane; Glycoprotein;
KW Signal.
FT SIGNAL. 1 21
FT CHAIN 22 173
FT FT T-CELL SURFACE GLYCOPROTEIN CD3 DELTA
FT FT CHAIN.

```

```

FT DOMAIN 22 100
FT TRANSMEM 101 127
FT POTENTIAL.
FT DOMAIN 128 173
FT CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 38 38
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 55 55
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 74 74
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 173 AA; 19356 MW; C462DF34E7FDB60 CRC64;

Query Match 88.6%; Score 39; DB 1; Length 173;
Best Local Similarity 70.0%; Pred. No. 0.57;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10
| | | | | | | |
Db 107 VIITDLIATL 116

RESULT 3
CD3D_SHEEP CD3D_SHEEP STANDARD; PRT; 167 AA.
AC P18438;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE T-CELL SURFACE GLYCOPROTEIN CD3 DELTA CHAIN PRECURSOR (T-CELL RECEPTOR
DE T3 DELTA CHAIN).
GN CD3D.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Lymphocytes;
RX MEDLINE=90353383; PubMed=2143727;
RA Hein W.R., Tunncliffe A.;
RA "Characterization of the CD3 gamma and delta invariant subunits of
RA the sheep T cell antigen receptor.";
RL Eur. J. Immunol. 20:1503-1511(1990).
CC -!- FUNCTION: THE CD3 COMPLEX MEDIATES SIGNAL TRANSDUCTION.
CC A TCR ALPHA/BETA OR TCR GAMMA/DELTA HETERODIMER COEXPRESSED AT THE
CC CELL SURFACE WITH THE INVARIANT SUBUNITS OF CD3 LABELED GAMMA,
CC DELTA, EPSILON, ZETA, AND ETA.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE GAMMA,
CC DELTA, AND EPSILON CHAINS ARE HIGHLY SIMILAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52993; CAA37182.1; -
CC PIR: S11725; RWSHD3.
CC PIR: B43547; B43547.
CC Immunoglobulin domain; T-cell; Receptor; Transmembrane; Glycoprotein;
KW Signal.
FT SIGNAL. 1 21
FT CHAIN 22 167
FT FT T-CELL SURFACE GLYCOPROTEIN CD3 DELTA
FT FT CHAIN.
FT FT
FT DOMAIN 22 96
FT TRANSMEM 97 123
FT POTENTIAL.
FT DOMAIN 124 167
FT CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 38 38
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 55 55
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 167 AA; 18562 MW; 35D55493AE369714 CRC64;

```



Query Match 84.1%; Score 37; DB 1; Length 167;  
 Best Local Similarity 60.0%; Pred. NO. 1.4;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIIVTDVIATL 10  
 DB 103 LIITDIATV 112

RESULT 4  
 APTL\_YEAST STANDARD; PRT; 187 AA.  
 AC P49435;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ADENINE PHOSPHORIBOSYLTRANSFERASE 1 (EC 2.4.2.7) (APRT 1).  
 GN APTL OR YML022W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DS1;  
 RX MEDLINE=95369738; PubMed=7642142;  
 RA Alfonso J.D., Sahota A., Deeley M.C., Ranjekar P., Taylor M.W.;  
 RT "Cloning and characterization of the adenine  
 phosphoribosyltransferase-encoding gene (APTL) from Saccharomyces  
 cerevisiae.";  
 RT Gene 161:81-85(1995).  
 RL [2]  
 RP REVISIONS.  
 RA Taylor M.W., Alfonso J.D.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CATALYSES A SALVAGE REACTION RESULTING IN THE FORMATION  
 OF AMP. THAT IS ENERGETICALLY LESS COSTLY THAN DE NOVO SYNTHESIS.  
 CC -!- CATALYTIC ACTIVITY: AMP + PYROPHOSPHATE -> ADENINE + 5-PHOSPHO-  
 CC ALPHA-D-RIBOSE 1-DIPHOSPHATE.  
 CC -!- PATHWAY: PURINE SALVAGE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE  
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; U16781; AAA89075.1; -;  
 EMBL; Z46659; CAA86633.1; -;  
 SGD; S0004484; APT1.  
 InterPro; IPR000836; -;  
 InterPro; IPR002375; -;  
 Pfam; PF00156; Pribosyltran; 1.  
 DR PROSITE; PS00103; PUR\_PYR\_PR\_TRANSFER; 1.  
 KW Transferase; Glycosyltransferase; Purine salvage; Multigene family.  
 SQ SEQUENCE 187 AA; 20637 MW; 0DC76B4BBD2395E2 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 187;  
 Best Local Similarity 77.8%; Pred. NO. 10;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIIVTDVIAT 9  
 DB 126 IIIVDDIIAT 134

RESULT 5  
 YK13\_CAEEL STANDARD; PRT; 398 AA.  
 ID YK13\_CAEEL  
 AC P34337;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL 44.9 KDA PROTEIN C15H7.3 IN CHROMOSOME III.  
 GN C15H7.3.  
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
 Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).

CC -!- SIMILARITY: SOME, TO PROTEIN-TYROSINE PHOSPHATASES.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; Z22173; CAA80125.1; -;  
 PIR; S40752; S40752;  
 WormPep; C15H7.3; CE000081.  
 DR InterPro; IPR000387; -;  
 DR InterPro; IPR000387; -;  
 DR Pfam; PF00102; Y\_Phosphatase; 1.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; FALSE\_NEG.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 KW Hypothetical protein; Hydrolase.  
 SQ SEQUENCE 398 AA; 44926 MW; CEDE09FA0BC8F68 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 398;  
 Best Local Similarity 55.6%; Pred. NO. 23;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 IVTDVIATL 10  
 DB 295 VTDLVATM 303

RESULT 6  
 UL06\_EBV STANDARD; PRT; 613 AA.  
 ID UL06\_EBV  
 AC P03213;  
 DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE VIRION PROTEIN BBRL.  
 GN BBRL.  
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 OC NCBI\_TaxID=10377;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
 RA Tuffnell P.S., Barrell B.G.;  
 RT "Data sequence and expression of the B95-8 Epstein-Barr virus genome."  
 RL Nature 310:207-211(1984)  
 CC -!- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA  
 CC PACKAGING (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,  
 CC EBV-1 56, EBV BBRL, HCMV UL104, AND VZV 54.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; V01555; CAA24820.1; -;  
 DR PIR; A03775; Q0BE33.  
 DR PIR; S33025; S33025.  
 DR InterPro; IPR002660; -;  
 DR Pfam; PF01763; Herpes\_UL6; 1.  
 KW Late protein.  
 SQ SEQUENCE 613 AA; 68456 MW; E6E65BB078FED9AD CRC64;  
  
 Query Match 75.0%; Score 33; DB 1; Length 613;  
 Best Local Similarity 77.8%; Pred. No. 36;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 2 IVTDVIATL 10  
 :|||||  
 Db 538 LVTVDVIATL 546  
  
 RESULT 7  
 PMAL\_AJECA  
 ID PMAL\_AJECA STANDARD; PRT; 916 AA.  
 AC Q07421;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).  
 GN PMAL.  
 OS Ajellomyces capsulata (Histoplasma capsulatum).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Onygenales; Onygenaceae; Ajellomyces.  
 OC NCBI\_TaxID=5037;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94124018; PubMed=7916725;  
 RA Schafer M.P., Dean G.E.;  
 RT "Cloning and sequence analysis of an H(+) ATPase-encoding gene from  
 RT the human dimorphic pathogen Histoplasma capsulatum."  
 RL Gene 136:295-300(1993).  
 CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A  
 CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE  
 CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING  
 CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALIZATION MAY MEDATE  
 CC GROWTH RESPONSES.  
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (E1-E2 ATPASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L07305; AAB53772.1; -;  
 DR InterPro; IPR000695; -;  
 DR InterPro; IPR001757; -;  
 DR Pfam; PF00122; E1-E2 ATPase; 1.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PRINTS; PR00120; HATPASE.  
 DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
 KW Hydrolase; Hydrogen ion transport; Transmembrane; Phosphorylation;  
 KW ATP-binding.  
 FT DOMAIN 1 111 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 112 134 1 (POTENTIAL).  
 FT DOMAIN 135 136 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 137 155 2 (POTENTIAL).  
 FT DOMAIN 156 279 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 280 305 3 (POTENTIAL).  
 FT DOMAIN 306 320 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 321 350 4 (POTENTIAL).  
 FT DOMAIN 351 711 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 712 734 5 (POTENTIAL).  
 FT DOMAIN 735 754 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 755 776 6 (POTENTIAL).  
 FT DOMAIN 777 817 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 818 843 7 (POTENTIAL).  
 FT DOMAIN 844 847 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 848 874 8 (POTENTIAL).  
 FT DOMAIN 875 916 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 933 962 ASP/GLU-RICH (ACIDIC).  
 FT MOD\_RES 374 374 PHOSPHORYLATION (BY SIMILARITY).  
 FT BINDING 470 470 ATP (BY SIMILARITY).  
 SQ SEQUENCE 916 AA; 98884 MW; 1B750ACE83F330AD CRC64;  
  
 Query Match 75.0%; Score 33; DB 1; Length 916;  
 Best Local Similarity 60.0%; Pred. No. 54;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 IIVTDVIATL 10  
 :|::|:|  
 Db 827 ILLVDIIATL 836  
  
 RESULT 8  
 PMAL2\_SCHPO  
 ID PMAL2\_SCHPO STANDARD; PRT; 1010 AA.  
 AC F28876;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PLASMA MEMBRANE ATPASE 2 (EC 3.6.1.35) (PROTON PUMP 2).  
 GN PMAL2 OR SPCC1393.01.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92011563; PubMed=18333395;  
 RA Ghisla M., Goffeau A.;  
 RT "The pmal and pmal2 H(+)-ATPases from Schizosaccharomyces pombe are  
 RT functionally interchangeable."

J. Biol. Chem. 266:18276-18279(1991).

[2]  
SEQUENCE OF 466-1010 FROM N.A.

STRAIN-972;  
RA Wood V., Rajandream M.A., Barrell B.G., Volkert G.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases  
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A  
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE  
CC ACTIVE TRANSPORT OF NUTRIENTS BY H<sup>+</sup>-SYMPORT. THE RESULTING  
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALIZATION MAY MEDIATE  
CC GROWTH RESPONSES.  
CC -!- CATALYTIC ACTIVITY: ATP + H<sub>2</sub>O = ADP + ORTHOPHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- PTM: IN ADDITION TO TRANSIENT PHOSPHORYLATION OF THE ACTIVE SITE  
CC ASP RESIDUE, THIS PROTEIN, BUT NOT THE PRODUCT OF THE PMAL LOCUS,  
CC IS PHOSPHORYLATED EFFICIENTLY IN ISOLATED PLASMA MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
CC (E1-E2 ATPASES).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M60471; AAA35325.1; ..  
EMBL; AL035592; CAB38157.1; ..  
DR PIR; A40945; PXZP2P.  
DR InterPro; IPR000695; ..  
DR InterPro; IPR001757; ..  
DR Pfam; PF00122; E1-E2ATPase; 1.  
DR PRINTS; PR00119; CATATPASE.  
DR PRINTS; PR00120; HATPASE.  
DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
KW Hydrolyase; Hydrogen ion transport; Transmembrane; Phosphorylation;  
KW ATP-binding; Multigene family.  
FT DOMAIN 1 201 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 202 222 1 (POTENTIAL).  
FT DOMAIN 223 226 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 227 246 2 (POTENTIAL).  
FT DOMAIN 247 377 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 378 399 3 (POTENTIAL).  
FT DOMAIN 400 410 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 411 433 4 (POTENTIAL).  
FT DOMAIN 434 805 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 806 824 5 (POTENTIAL).  
FT DOMAIN 825 840 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 841 860 6 (POTENTIAL).  
FT DOMAIN 861 912 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 913 933 7 (POTENTIAL).  
FT DOMAIN 934 946 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 947 963 8 (POTENTIAL).  
FT DOMAIN 964 1010 CYTOPLASMIC (POTENTIAL).  
FT MOD\_RES 464 464 PHOSPHORYLATION (PROBABLE).  
FT BINDING 560 560 ATP (BY SIMILARITY).  
FT DOMAIN 121 150 ASP/GLU-RICH (ACIDIC).  
SQ SEQUENCE 1010 AA; 110127 MW; 2C629A45125B4DC3 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 1010;  
Best Local Similarity 50.0%; Pred. No. 60;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10  
:::|::|::|  
Db 917 VLVVDILATL 926

RESULT 9  
NAGA\_CAEEL STANDARD; PRT; 418 AA.  
ID NAGA\_CAEEL

AC P34480;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PUTATIVE N-ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE (EC 3.5.1.25)  
DE (GLCNAC 6-P DEACETYLASE).  
DE F59B2.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almscoog R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Ropra A., Saunders D., Showkeen R.,  
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
CC -!- CATALYTIC ACTIVITY: N-ACETYL-D-GLUCOSAMINE 6-PHOSPHATE + H<sub>2</sub>O  
CC -!- D-GLUCOSAMINE 6-PHOSPHATE + ACETATE.  
CC -!- SIMILARITY: BELONGS TO THE NAGA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Z11505; CAA77585.1; ..  
DR PIR; S31124; S31124.  
DR WormPep; F59B2.3; CE00231.  
KW Hypothetical protein; Hydrolase; Carbohydrate metabolism.  
SQ SEQUENCE 418 AA; 44887 MW; 4988EA416F5D1247 CRC64;

Query Match 72.7%; Score 32; DB 1; Length 418;  
Best Local Similarity 60.0%; Pred. No. 39;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10  
:::|::|::|  
Db 302 VLVTDIAAL 311

RESULT 10  
PMAL\_KLUJA STANDARD; PRT; 899 AA.  
ID PMAL\_KLUJA  
AC P49380;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).  
GN PMAL.  
OS Kluyveromyces lactis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
OX NCBI\_TaxID=28985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WM37;

RX MEDLINE=95247668; PubMed=7730265;  
RA Miranda M., Ramirez J., Pena A., Coria R.;  
RT "Molecular cloning of the plasma membrane H(+)-ATPase from  
RT Kluyveromyces fragilis: a single nucleotide substitution in the gene  
RT confers ethidium bromide resistance and deficiency in K+ uptake.";  
RT J. Bacteriol. 177:2360-2367(1995).  
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A  
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE  
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+ SYMPORT. THE RESULTING  
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALINIZATION MAY MEDIATE  
CC GROWTH RESPONSES.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.  
CC -!- ENZYME REGULATION: ACTIVATED BY HIGH PH OR ALSO BY POTASSIUM IONS  
CC WHEN THE MEDIUM PH IS LOW.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
CC (E1-E2 ATPASES).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: L37875; AAA69688.1; -;  
CC InterPro: IPR000695; -;  
CC InterPro: IPR001757; -;  
CC Pfam: PF00122; E1-E2\_ATPase; 1.  
CC PRINTS: PR00119; CATATPASE.  
CC PRINTS: PR00120; HATPASE.  
CC PROSITE: PS00154; ATPASE\_E1\_E2; 1.  
CC Hydrolase; Hydrogen ion transport; Transmembrane; Phosphorylation;  
CC ATP-binding.  
CC  
CC DOMAIN 1 96 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 97 117 1 (POTENTIAL).  
CC DOMAIN 118 121 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 122 141 2 (POTENTIAL).  
CC DOMAIN 142 272 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 273 294 3 (POTENTIAL).  
CC DOMAIN 295 305 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 306 328 4 (POTENTIAL).  
CC DOMAIN 329 700 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 701 719 5 (POTENTIAL).  
CC DOMAIN 720 735 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 736 755 6 (POTENTIAL).  
CC DOMAIN 756 805 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 806 826 7 (POTENTIAL).  
CC DOMAIN 827 838 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 839 855 8 (POTENTIAL).  
CC DOMAIN 856 899 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 899 959 PHOSPHORYLATION (PROBABLE).  
CC MOD\_RES 359 455 ATP (BY SIMILARITY).  
CC BINDING 455 571 POLY-GLY.  
CC DOMAIN 566 669 M -> I (IN MUTANT 3.3; LOW CAPACITY TO  
CC VARIANT PUMP OUT PROTONS).  
CC  
CC SEQUENCE 899 AA; 98259 MW; F29DC853BDCF4396 CRC64;  
  
CC Query Match 72.7%; Score 32; DB 1; Length 899;  
CC Best Local Similarity 40.0%; Pred. No. 85;  
CC Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PHOSPHOLIPID-TRANSPORTING ATPASE 1 (EC 3.6.3.13).  
GN ALA1 OR AT5G04930 OR MUG13.22.  
OS Arabidopsis thaliana (Mouse-ear cress).  
CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II;  
CC Brassicales; Brassicaceae; Arabidopsis.  
CC NCBI\_TaxID=3702;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=97471969; PubMed=9330910;  
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
RT features of the 1.6 Mb regions covered by twenty physically assigned  
RT pl clones.";  
RL DNA Res. 4:215-230(1997).  
CC -!- FUNCTION: INVOLVED IN TRANSPORT OF PHOSPHOLIPIDS.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
CC (E1-E2 ATPASES). SUBFAMILY IV.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: AB005245; BAB11515.1; -;  
CC PROSITE: PS00154; ATPASE\_E1\_E2; 1.  
CC Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Metal-binding;  
CC Magnesium; Multigene family.  
CC DOMAIN 1 100 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 101 122 POTENTIAL.  
CC DOMAIN 123 127 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 128 150 POTENTIAL.  
CC DOMAIN 151 329 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 330 351 POTENTIAL.  
CC DOMAIN 352 391 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 392 409 POTENTIAL.  
CC DOMAIN 410 914 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 915 934 POTENTIAL.  
CC DOMAIN 935 948 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 949 968 POTENTIAL.  
CC DOMAIN 969 998 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 999 1020 POTENTIAL.  
CC DOMAIN 1021 1027 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 1028 1050 POTENTIAL.  
CC DOMAIN 1051 1056 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 1057 1077 POTENTIAL.  
CC DOMAIN 1078 1090 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 1091 1115 POTENTIAL.  
CC DOMAIN 1116 1158 CYTOPLASMIC (POTENTIAL).  
CC MOD\_RES 457 457 PHOSPHORYLATION (BY SIMILARITY).  
CC METAL 859 859 MAGNESIUM (BY SIMILARITY).  
CC METAL 863 863 MAGNESIUM (BY SIMILARITY).  
CC SEQUENCE 1158 AA; 130328 MW; 5CC042B40C8C974D CRC64;

Query Match 72.7%; Score 32; DB 1; Length 1158;  
Best Local Similarity 70.0%; Pred. No. 11e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 IIVTDVIATL 10  
Db 1073 VIVIDVIATL 1082

RESULT 11  
ID ALAL\_ATHATH STANDARD; PRT; 1158 AA.  
AC P98204;

```

RESULT 12
TPFL_TREPA
ID TPFL_TREPA STANDARD; PRT; 176 AA.
AC P16665;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ANTIGEN TPFL (ANTIGEN C1-5) (ANTIGEN 4D).
GN TPFL OR TP1038.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC STRAIN=NICHOLS;
RX MEDLINE=89281138; PubMed=2471912;
RA Noordhoek G.T., Hermans P.W.M., Paul A.N., Schouls L.M.,
RA van der Sluis J.J., van Embden J.D.A.;
RT "Treponema pallidum subspecies pallidum (Nichols) and Treponema
RT pallidum subspecies pertenuis (CDC 2575) differ in at least one
RT nucleotide: comparison of two homologous antigens.";
RL Microb. Pathog. 6:29-42(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89108612; PubMed=2643579;
RA Walfield A.M., Roche E.S., Zounes M.C., Kirkpatrick H., Wild M.A.,
RA Textor G., Tsai P.K., Richardson C.;
RT "Primary structure of an oligomeric antigen of Treponema pallidum.";
RL Infect. Immun. 57:633-635(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
RN [4]
RP DISULFIDE BONDS.
RC STRAIN=NICHOLS;
RX MEDLINE=87165736; PubMed=3549683;
RA Radolf J.D., Borenstein L.A., Kim J.Y., Fehniger T.E., Lovett M.A.;
RT "Role of disulfide bonds in the oligomeric structure and protease
RT resistance of recombinant and native Treponema pallidum surface
RT antigen 4b.";
RL J. Bacteriol. 169:1365-1371(1987).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT STRUCTURAL ROLE IN THE OUTER
CC MEMBRANE.
CC -1- SUBUNIT: HOMODECAMER; EITHER LINKED OR STABILIZED BY DISULFIDE
CC BONDS.
CC -1- SIMILARITY: BELONGS TO THE DPS FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M32401; AAC26592.1;
DR APT_MYCGE
ID APT_MYCGE STANDARD; PRT; 180 AA.
AC P47518;
DR TIGR; TP1038;
DR InterPro; IPR002177;
DR Pfam; PF02047; DPS; 1.
Query Match 70.5%; Score 31; DB 1; Length 176;
Best Local Similarity 55.6%; Pred. No. 25;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 IYTDVIATL 10
DB 155 IITDILRTL 163
-----
RESULT 14
APT_MYCGE
ID APT_MYCGE STANDARD; PRT; 180 AA.
AC P47518;
DR TIGR; TP1038;
DR InterPro; IPR002177;
DR Pfam; PF02047; DPS; 1.
Query Match 70.5%; Score 31; DB 1; Length 176;
Best Local Similarity 55.6%; Pred. No. 25;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 IYTDVIATL 10
DB 155 IITDILRTL 163
-----
RESULT 16
APT_MYCGE
ID APT_MYCGE STANDARD; PRT; 180 AA.
AC P47518;
DR TIGR; TP1038;
DR InterPro; IPR002177;
DR Pfam; PF02047; DPS; 1.
Query Match 70.5%; Score 31; DB 1; Length 176;
Best Local Similarity 55.6%; Pred. No. 25;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 IYTDVIATL 10
DB 155 IITDILRTL 163
-----

```

```

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.7) (APRT).
GN APT OR MG276.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC 33530 / G-37;
RC MEDLINE=96026346; PubMed=7569993;
RX Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.B., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RN SEQUENCE OF 144-180 FROM N.A.
RP STRAIN-ATCC 33530 / G-37;
RC MEDLINE=94075230; PubMed=8253680;
RX Peterson S.N., Hu P.-C., Boff K.F., Hutchison C.A. III;
RA "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- FUNCTION: CATALYSES A SALVAGE REACTION RESULTING IN THE FORMATION
CC OF AMP, THAT IS ENERGETICALLY LESS COSTLY THAN DE NOVO SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: AMP + PYROPHOSPHATE = ADENINE + 5-PHOSPHO-
CC ALPHA-D-RIBOSE 1-DIPHOSPHATE.
CC -!- PATHWAY: PURINE SALVAGE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
CC PHOSPHORIBOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39707; AAC71498.1; -;
DR EMBL; U01786; AAD10608.1; -;
DR TIGR; MG276; -;
DR InterPro; IPR000836; -;
DR InterPro; IPR002375; -;
DR Pfam; PF00156; Pribosyltran. 1.
DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER. 1.
KW Transferase; Glycosyltransferase; Purine salvage.
SQ SEQUENCE 180 AA; 19976 MW; 81E63CDB8DF1F234 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 180;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IIIVTDVIAT 9
Db 124 VIVDDVLAT 132
||| |||

RESULT 15
IL2A_SHEEP STANDARD; PRT; 275 AA.
AC P26898;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INTERLEUKIN-2 RECEPTOR ALPHA CHAIN PRECURSOR (IL-2 RECEPTOR ALPHA

DE SUBUNIT) (P55) (TAC ANTIGEN) (CD25) .
GN IL2RA.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=T-cell;
RC Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RN MEDLINE=92241682; PubMed=1572550;
RX Bujdosó R., Sargan D.R., Williamson M.L., McConnell I.;
RA "Cloning of a cDNA encoding the ovine interleukin-2 receptor 55-kDa
RT protein, CD25.";
RL Gene 113:283-284(1992).
CC -!- FUNCTION: RECEPTOR FOR INTERLEUKIN-2.
CC -!- SUBUNIT: NON COVALENT DIMER OF AN ALPHA AND A BETA CHAINS. IL-2R
CC EXIST IN 3 DIFFERENT FORMS: A HIGH AFFINITY DIMER, AN INTERMEDIATE
CC AFFINITY MONOMER (BETA CHAIN), AND A LOW AFFINITY MONOMER (ALPHA
CC CHAIN). THE HIGH AND INTERMEDIATE AFFINITY FORMS ALSO ASSOCIATE
CC WITH A GAMMA CHAIN.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11560; CAA77652.1; -;
DR EMBL; X60149; CAA42723.1; -;
DR EMBL; A19167; CAA01447.1; -;
DR PIR; S18910; S18910.
DR PIR; S18899; S18899.
DR PIR; JG1113; JG1113.
DR HSP; P01589; IILN.
DR InterPro; IPR000436; -;
DR Pfam; PF00084; sushi; 2.
KW Transmembrane; Glycoprotein; Receptor; Repeat; Signal; Sushi.
FT SIGNAL 1 21
FT CHAIN 22 275
FT DOMAIN 22 243
FT TRANSMEM 244 262
FT DOMAIN 263 275
FT DOMAIN 23 78
FT DOMAIN 122 185
FT DISULFID 24 64
FT DISULFID 251 77
FT DISULFID 123 168
FT DISULFID 152 184
FT CARBOHYD 80 80
FT CONFLICT 166 166 S -> T (IN REF. 2).
SQ SEQUENCE 275 AA; 30904 MW; 1101A2DE5AC5A088 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 275;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 TDVIAIL 10
Db 227 TDVVATL 233
||| |||

Search completed: May 10, 2001, 10:15:24

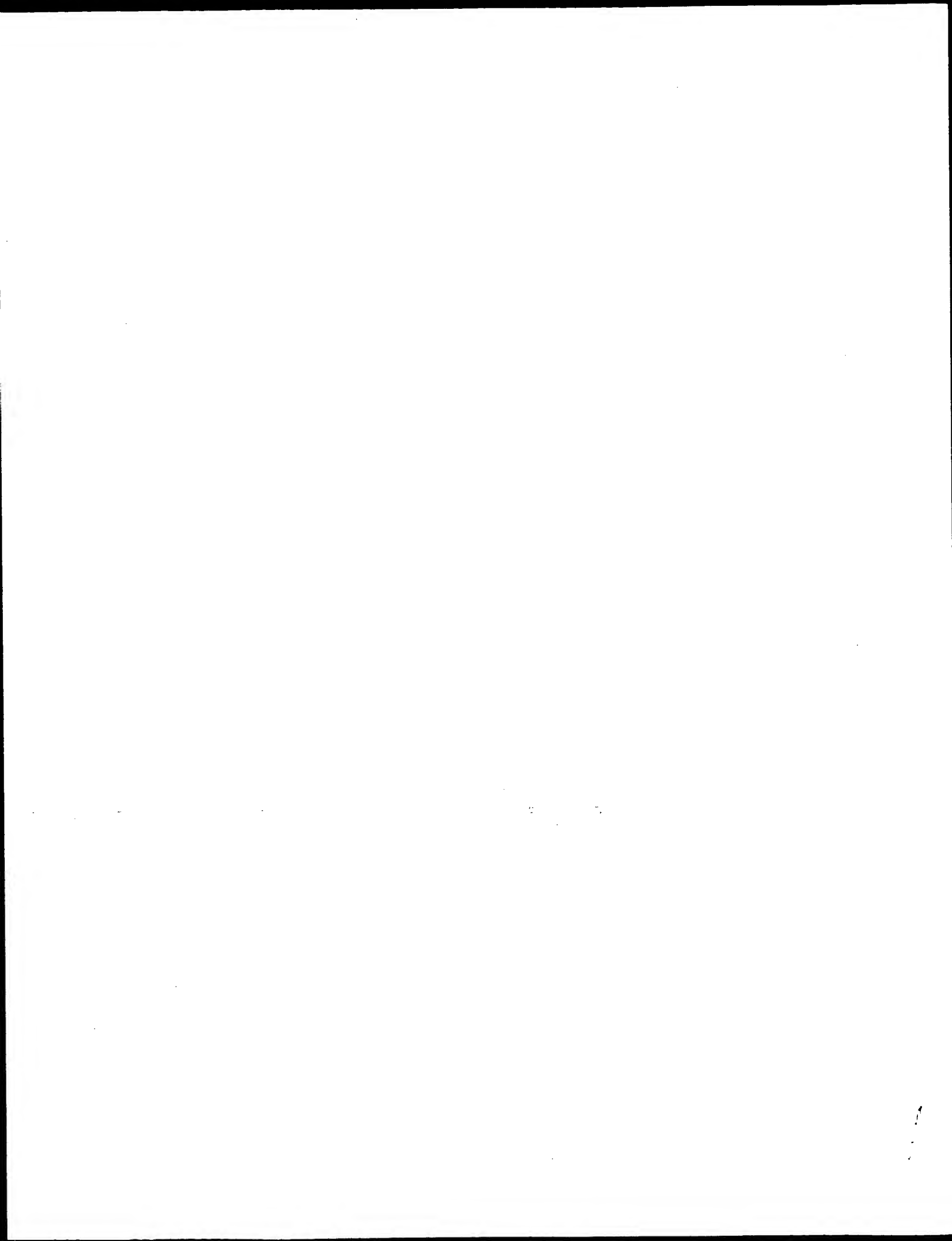
```

Fri May 11 08:35:17 2001

us-09-202-305-13.rsp

Page 9

Job time: 253 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 10:10:33 ; Search time 51.82 Seconds  
(without alignments)  
11.031 Million cell updates/sec

Title: US-09-202-305-12

Perfect score: 50

Sequence: 1. LMTLRLWSS 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	10	19 W42363	TCR-alpha intracel
2	50	100.0	35	18 W33992	Human T-cell antig
3	50	100.0	266	20 Y05727	Mouse A1 T cell re
4	50	100.0	267	17 W04300	Murine T-cell rece
5	50	100.0	267	19 W47588	T-cell receptor al
6	50	100.0	268	7 P60237	Sequence of the al
7	50	100.0	268	16 R77288	T-cell receptor al
8	50	100.0	268	19 W36110	Mouse T-cell recep
9	50	100.0	269	21 P50995	Human receptor-ass
10	50	100.0	270	6 P50256	Sequence of T-cell
11	50	100.0	272	19 W36111	Mouse T-cell recep

12	50	100.0	273	20 Y05405	Killer T-cell rece
13	50	100.0	274	20 Y05404	Killer T-cell rece
14	50	100.0	277	7 P60065	Sequence of a poly
15	42	84.0	99	22 B63434	Human breast cance
16	37	74.0	174	7 P60435	Sequence of a port
17	37	74.0	174	14 R44423	Peptide coded by p
18	37	74.0	174	15 P58710	Human PIP fragment
19	37	74.0	174	18 W31168	hPIP partial fragm
20	37	74.0	335	7 P60436	Sequence of human
21	37	74.0	335	11 R07561	Recombinant urokin
22	37	74.0	335	14 R44424	Human phospholipas
23	37	74.0	335	15 P58707	Human phospholipas
24	37	74.0	335	17 R97612	Human urokinase pl
25	37	74.0	335	18 W31165	Human phospholipas
26	37	74.0	335	20 Y04103	Urokinase-type pla
27	37	74.0	335	21 B26241	Human urokinase-ty
28	35	70.0	106	20 Y37127	Chlamydia trachoma
29	34	68.0	28	19 W63021	Mouse dectin-2 tra
30	34	68.0	134	19 W63019	Mouse dectin-2 iso
31	34	68.0	168	19 W63020	Mouse dectin-2 iso
32	34	68.0	175	19 W63018	Mouse dectin-2 iso
33	34	68.0	200	21 G22358	Arabidopsis thalia
34	34	68.0	200	21 G60143	Arabidopsis thalia
35	34	68.0	209	19 W63010	Mouse dectin-2. M
36	34	68.0	209	20 Y27447	Mouse SDCMP3 poly
37	34	68.0	211	21 B40453	Human ORFX ORF217
38	34	68.0	289	21 G09540	Arabidopsis thalia
39	34	68.0	289	21 G43892	Arabidopsis thalia
40	34	68.0	291	14 R34785	KREL. Saccharomyc
41	34	68.0	299	21 G09539	Arabidopsis thalia
42	34	68.0	299	21 G43891	Arabidopsis thalia
43	34	68.0	309	21 G09538	Arabidopsis thalia
44	34	68.0	309	21 G43890	Arabidopsis thalia
45	34	68.0	324	21 G22357	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

ID W42363 standard; peptide; 10 AA.

XX AC W42363;

DT 11-JUN-1998 (first entry)

XX TCR-alpha intracellular chain derived peptide 1.

DE Peptide; TCR-alpha intracellular chain; T-cell receptor; inflammation;  
KW autoimmune disease; inflammatory bowel disease; psoriasis; infection;  
KW acquired immune deficiency syndrome; allergy.

XX OS Synthetic.

OS Homo sapiens.

XX W09747644-A1.

XX PD 18-DEC-1997.

XX PF 11-JUN-1997; 97WO-AU00367.

XX PR 11-JUN-1996; 96AU-0000394.

XX PR 11-JUN-1996; 96AU-0000389.

XX PR 11-JUN-1996; 96AU-0000390.

XX PR 11-JUN-1996; 96AU-0000391.

XX PR 11-JUN-1996; 96AU-0000392.

XX PR 11-JUN-1996; 96AU-0000393.

XX PA (NSYD-) NORTHERN SYDNEY AREA HEALTH SERVICE.

XX PI Manolios N;

XX

DR WPI: 1998-052238/05.

XX New peptide(s) that inhibit the T cell receptor - used to treat  
PT inflammation, autoimmune disease, allergy etc. and to deliver  
PT conjugated therapeutic agents to cells

XX Claim 5; Page 31; 58pp; English.

XX The present peptide sequence is derived from the T-cell receptor  
CC (TCR)-alpha intracellular chain. This peptide and others  
CC (see W42357-W42371) act to inhibit TCR function probably by interfering  
CC with TCR assembly. They are claimed to be useful in treating disorders  
CC in which T cells are involved or recruited, e.g. allergy, autoimmune  
CC diseases, inflammatory bowel disease, psoriasis, infections (including  
CC acquired immune deficiency syndrome) etc. They are also claimed to be  
CC able to deliver conjugated therapeutic agents to cells.

XX Sequence 10 AA;

Query Match 100.0%; Score 50; DB 19; Length 10;  
Best Local Similarity 100.0%; Pred. NO. 0.0035;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LLMTLRLWSS 10  
DB 1 llmtlrlwss 10  
|||||

RESULT 2

ID W33992 standard; peptide; 35 AA.

XX AC W33992;

XX 28-MAY-1998 (first entry)

XX Human T-cell antigen receptor antagonist.

XX Antagonist; integral membrane protein; inhibitor; transmembrane domain;  
KW therapy; psychotic disorder; Huntington's disease; Parkinson's disease;  
KW post-myocardial infarction; HIV infection; cell proliferation disorder;  
KW peptic ulcer; glaucoma; asthma; renal disease; autoimmune disease;  
KW antibacterial; human T-cell antigen.

XX Homo sapiens.

XX WO9735881-A2.

XX 02-OCT-1997.

XX 26-MAR-1997; 97WO-CA00203.

XX 20-AUG-1996; 96US-0024240.

XX 27-MAR-1996; 96US-0014306.

XX 25-JUN-1996; 96US-0670119.

XX (GEOR/) GEORGE S R.

XX (NGGY/) NG G Y K.

XX (ODOW/) O'DOWD B F.

XX (SEEM/) SEEMAN P.

XX George SR, Ng GYK, O'Dowd BF, Seeman P;

XX WPI: 1997-489566/45.

XX Inhibitor of integral membrane protein - used to treat e.g.

XX schizophrenia, hypertension, viral infection, cancer etc.

XX Claim 35; Page 98; 127pp; English.

XX This sequence is a human T-cell antigen receptor antagonist, and is an  
CC antagonist of the invention. The antagonists (A) are for inhibiting the

CC function of a prokaryotic or eukaryotic integral membrane protein (IMP)  
CC having at least one transmembrane domain (TMD), comprises a peptide  
CC including at least 4 consecutive amino acids (aa) from the sequence of  
CC the IMP. (A) are used to treat or prevent disorders in mammals that  
CC involve disturbances of IMP, and the same effect is achieved by inserting  
CC a nucleic acid encoding (A) into the cells of the mammal. Specified  
CC conditions that can be treated are schizophrenia, psychotic disorders,  
CC Huntington's disease, Tourette's syndrome, hypertension, post-myocardial  
CC infarction, tachyarrhythmia, human immunodeficiency virus infection,  
CC AIDS, cell proliferation (cancer, psoriasis or hyperkeratosis), and drug  
CC abuse. (A) can also be used to treat peptic ulcer, glaucoma, asthma,  
CC renal disease, Parkinson's disease, congestive heart failure, migraine,  
CC depression, obesity, diabetic neuropathy/retinopathy, autoimmune  
CC diseases, graft rejection etc. (A) can be used as diuretics and  
CC antibacterials. (A) may also be used to determine function of  
CC orphan receptors and, when attached to a marker, for tissue imaging to  
CC localise/quantify specific receptors. The transgenic animals are used to  
CC test efficacy/toxicity of (A) and as models of diseases. (A) are very  
CC specific and selective, against the IMP from which they are derived, so  
CC should have few side effects.

XX Sequence 35 AA;

Query Match 100.0%; Score 50; DB 18; Length 35;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LLMTLRLWSS 10  
DB 26 llmtlrlwss 35  
|||||

RESULT 3

ID Y05727 standard; Protein; 266 AA.

XX AC Y05727;

XX 19-JUL-1999 (first entry)

XX Mouse Al T cell receptor alpha chain.

XX T cell receptor alpha chain; mouse; transgenic animal;  
KW animal model; immunological tolerance; graft rejection;  
KW tissue grafting; neonatal intolerance; transplantation antigen;  
KW H-Y antigen.

XX Mus musculus.

XX WO9916867-A1.

XX 08-APR-1999.

XX 30-SEP-1998; 98WO-GB02965.

XX 01-OCT-1997; 97GB-0020888.

XX (ISIS-) ISIS INNOVATION LTD.

XX Cobbold SP, Waldmann H, Zelenika D;

XX WPI: 1999-255090/21.

XX N-PSDB; X25357.

XX Transgenic non-human mammal having only CD4 positive T cells  
PT specific for at least one transplantation antigen, useful for  
PT studying immunological tolerance

XX Disclosure; Page 32-33; 41pp; English.

XX The present sequence represents the T cell receptor (TCR) alpha  
CC chain from the Al CD4+ T cell clone isolated from CBA/Ca mice. The

CC A1 clone recognises the minor histocompatibility antigen H-Y  
 CC present in male, but absent in female, mice. cDNA (see X25357)  
 CC encoding the A1 TCR alpha chain was obtained by RT-PCR (see  
 CC also X25350-51). and was used in the construction of A1(M)  
 CC transgenic mice. The invention relates to a transgenic animal  
 CC model comprising TCR alpha and beta chains. It provides a  
 CC genetically modified non-human mammal having a population of CD4  
 CC positive T cells specific for one or a limited number of selected  
 CC antigens, including at least transplantation antigen capable of  
 CC rejecting a tissue transplant containing the transplantation  
 CC antigen and, if applicable, the other selected antigens. The  
 CC animal has TCR genes which encode a TCR specific for the  
 CC transplantation antigen. The animal is useful for studying  
 CC immunological tolerance, especially the mechanisms of tolerance to,  
 CC and the rejection of, tissue grafts, and in pregnancy. The animals  
 CC are also useful for testing agents for biological activity in  
 CC promoting or reducing immunological tolerance.

XX Sequence 266 AA;

Query Match 100.0%; Score 50; DB 20; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRWSS 10  
 |||||  
 Db 257 llmtlrlwss 266

RESULT 4

W04300  
 ID W04300 standard; Protein: 267 AA.

XX AC W04300;

XX DT 02-JAN-1997 (first entry)

XX DE Murine T-cell receptor alpha chain.

XX KW Antigen-specific glycosylation inhibiting factor; AgGIF; lymphokine;  
 KW T-cell receptor; TCR; immunosuppressant; immunosuppressive;  
 KW hypersensitivity; graft rejection; autoimmune disease.

XX OS Mus sp.

XX FH Key Location/Qualifiers  
 XX FT Peptide 1..20  
 XX FT Region /label= Leader\_peptide  
 XX FT Region 21..109  
 XX FT Region /label= V-alpha\_region  
 XX FT Region 110..130  
 XX FT Region /label= J-alpha\_region  
 XX FT Region 131..267  
 XX FT Region /label= C-alpha\_region

XX PN W09631617-A1.

XX PD 10-OCT-1996.

XX PF 25-SEP-1995; 95WO-US12309.

XX PR 04-APR-1995; 95US-0416336.

XX PA (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.

XX PI Ishii Y, Ishizaka K;

XX DR WPI; 1996-465037/46.

XX DR N-PSDB; T38389.

XX PT Recombinant antigen-specific glycosylation-inhibiting factor -  
 XX useful as immunosuppressant

XX PS

XX Example 3; Fig 2A; 72pp; English.

CC A cDNA clone (T38389) coding for the T-cell receptor (TCR) alpha  
 CC chain (W04300) was isolated from T-cell hybridoma 231F1 cells  
 CC (ovalbumin-specific). TCR alpha plays a role in the formation  
 CC of antigen-specific glycosylation inhibiting factor (AgGIF) chains  
 CC which directly bind to an antigen and suppress the immune response  
 CC generated against that antigen; AgGIF is an expression product of  
 CC the TCR alpha chain gene (see also T38389). A polynucleotide  
 CC encoding AgGIF can be obt'd. from activated suppressor T-cells  
 CC using a probe based on non-specific GIF cDNA (see also T38390-91)  
 CC and a probe based on TCR alpha cDNA, and used to direct expression  
 CC of AgGIF in host cells.

XX Sequence 267 AA;

Query Match 100.0%; Score 50; DB 17; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRWSS 10  
 |||||  
 Db 258 llmtlrlwss 267

RESULT 5

W47588  
 ID W47588 standard; Protein: 267 AA.

XX AC W47588;

XX DT 26-JUN-1998 (first entry)

XX DE T-cell receptor alpha-chain.

XX KW Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;  
 KW prevention; therapy; tumour disease; renal cell carcinoma.

XX OS Homo sapiens.

XX PN DE19625191-A1.

XX PD 02-JAN-1998.

XX PF 24-JUN-1996; 96DE-1025191.

XX PR 24-JUN-1996; 96DE-1025191.

XX PA (BOEF ) BOEHRINGER MANNHEIM GMBH.

XX PI Schendel D;

XX DR WPI; 1998-053442/06.

XX DR N-PSDB; V18705.

XX PT Human T-cell receptor nucleic acids and poly-peptide(s) - for  
 XX diagnosis or therapy, especially of renal cell carcinoma

XX PS Example 1; Pages 11-13; 30pp; German.

XX CC The present sequence is the alpha-chain of a human T-cell  
 XX receptor (TCR), which can be used in the diagnosis, monitoring,  
 XX prevention and therapy of a tumour disease, specifically renal  
 XX cell carcinoma.

XX SQ Sequence 267 AA;

Query Match 100.0%; Score 50; DB 19; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PS Example; Fig 4a; 50pp; English.  
XX T-cell specific cDNA clones were isolated from an alloreactive CTL  
CC clone 2C of Balb/c mouse origin and specific for the D end of the  
CC Balb/C H-2 complex. A library of cDNA was constructed and screened  
CC with hybridisation probes. Two distinct classes of T-cell-specific  
CC cDNA clones whose genes are rearranged in cytotoxic T lymphocytes  
CC (CTLs) were identified (see N60166 and N60167). A heterodimeric  
CC T lymphocyte receptor comprising an alpha and a beta subunit is  
CC claimed.  
XX  
XX  
SQ Sequence 268 AA;

Query Match 100.0%; Score 50; DB 7; Length 268;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LLMTLRLWSS 10  
||| |||||  
DB 259 llmtlrlwss 268

RESULT 7  
R77288  
ID R77288 standard; Protein: 268 AA.  
XX  
AC R77288;  
XX  
DT 24-NOV-1995 (first entry)  
XX  
DE T-cell receptor alpha chain (TCR alpha) from hybridoma 3B3.  
XX  
KW T-cell receptor alpha chain: TCR alpha; hyperimmune;  
KW immunodeficiency; Cd4+; helper T cell; hybridoma 3B3.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..20  
FT /label= signal  
XX  
PN W09516462-A.  
XX  
XX 22-JUN-1995.  
XX  
PD 13-DEC-1994; 94WO-US14542.  
XX  
PF 13-DEC-1993; 93US-0165496.  
XX  
PR (KIRI ) KIRIN BEER KK.  
PA (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.  
XX  
XX Bissonnette R, Fotedar A, Green D, Ishii Y, Mikayama T;

PI WPI: 1995-231357/30.  
XX  
DR N-PSDB; Q91362.  
XX  
XX Antigen specific immune modulation using T cell receptor alpha chain -  
PT for treating allergy, cancer, auto-immune disease etc.  
XX  
XX Example; Figure 15; 119pp; English.  
XX  
XX TCR alpha cDNA of 3B3 cells was cloned by PCR. mRNA was isolated  
CC from 3B3 cells. cDNA was generated. cDNAs were ligated at the 5'  
CC and 3' ends using T4 ligase to construct circular DNA. Oligo  
CC primers encoding murine C alpha DNA were synthesised (see Q91363,  
CC Q91364). PCR was carried out. Amplified cDNA was subcloned into  
CC PCR1000 vector of TA cloning system. Three different TCR alpha  
CC cDNA were cloned and sequenced. Two of them were identified to be  
CC originated from the fusion partner cell of 3B3 hybridoma, BW5147.  
CC The other was confirmed not to be expressed in BW5147 which  
CC indicated that this TCR alpha originated from PLA2-specific T

QY 1 LLMTLRLWSS 10  
||| |||||  
DB 258 llmtlrlwss 267  
RESULT 6  
P60237  
ID P60237 standard; Protein: 268 AA.  
XX  
AC P60237;  
XX  
DT 19-AUG-1991 (first entry)  
XX  
DE Sequence of the alpha subunit of a clonally diversified  
DE integral membrane protein encoded by clone PHDS58.  
XX  
XX T-cell clone specific antibody; cytotoxic T lymphocyte; immunoassay;  
KW drug delivery.  
XX  
XX Mus musculus.  
OS

XX  
FH Key Location/Qualifiers  
FT Peptide 1..20  
FT /label= signal  
FT Region 21..118  
FT /label= variable  
FT Region 119..131  
FT /label= joining  
FT Region 132..241  
FT /label= constant  
FT Region 242..263  
FT /label= transmembrane  
FT Region 264..268  
FT /label= cytoplasmic  
FT Binding-site 90..92  
FT /label= potential N-glycosylation site  
FT Binding-site 198..200  
FT /label= see above  
FT Binding-site 212..214  
FT /label= see above  
FT Binding-site 241..243  
FT /label= see above  
FT Disulfide-bond 22..110  
FT Disulfide-bond 154..204  
FT Disulfide-bond 222  
FT /note= "inter-chain"

PN EP180878-A.  
XX  
XX 14-MAY-1986.  
XX  
XX 26-OCT-1985; 85EP-0113658.  
XX  
XX 31-OCT-1984; 84US-0666988.  
XX  
XX 13-JUN-1984; 84US-0620122.  
XX  
XX 01-MAR-1984; 84US-0585333.  
XX  
XX 22-OCT-1984; 84US-0663809.  
XX  
XX 27-JUL-1989; 89US-0385897.  
XX  
XX (MASI ) MASSACHUSETTS INST TECH.  
PA (STRD ) LELAND STANFORD JR UNIV.  
XX  
XX Saito H, Kranz DM, Eisen HN, Tonegawa S;  
XX  
XX WPI: 1986-126342/20.  
XX  
XX N-PSDB; N60166.

XX  
XX New hetero-dimeric t-lymphocyte receptor and its sub-units -  
PT useful in prodn. of T-cell clone specific antibodies for  
PT immunoassays, isolation procedures, specific delivery of bound  
PT drugs etc.  
XX

CC cells. Two of independent clones encoding this TCR alpha cDNA were  
 CC isolated and their DNA sequences were confirmed to be identical.  
 CC The DNA sequence of this 3B3 derived TCR alpha cDNA is shown in  
 CC Q91362 and its deduced AA sequence is in R77288.

XX SQ Sequence 268 AA;

Query Match 100.0%; Score 50; DB 16; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRWSS 10  
 Db 259 llmtlrwss 268  
 |||||

RESULT 8  
 W36110  
 ID W36110 standard; Protein; 268 AA.  
 XX AC W36110;

XX 19-MAY-1998 (first entry)

XX Mouse T-cell receptor alpha-chain protein from cell line 3B3.

XX Mouse; T-cell receptor; alpha-chain constant region; antigen-specific;  
 KW immunosuppressant; humoral; cell mediated immune response; allergy;  
 KW hypersensitivity; autoimmune reaction; transplant rejection.

XX Mus sp.

XX WO9743411-AL.

XX 20-NOV-1997.

XX 09-MAY-1997; 97WO-JP01565.

XX 29-MAY-1996; 96JP-0135572.

XX 10-MAY-1996; 96JP-0116101.

XX (KIRI ) KIRIN BEER KK.

XX Honma N, Mikayama T, Yuyama N;

XX WPI; 1998-008880/01.

XX N-PSDB; V01408.

XX Immunosuppressant peptide containing T-cell receptor alpha-chain  
 PT sequence - are not antigen-specific and do not induce antibody  
 PT production

XX Example 1; Page 37-38; 63pp; Japanese.

XX The present sequence represents mouse T-cell receptor alpha-chain  
 CC from cell line 3B3. The protein is an immunosuppressant which is not  
 CC antigen-specific and suppresses both humoral and cell-mediated immune  
 CC reactions. It can be used for treatment and/or prevention of delayed  
 CC hypersensitivity reactions, allergies and autoimmune reactions, and  
 CC inhibition of transplant rejection. The protein does not induce the  
 CC formation of antibodies against them to any significant extent.

XX SQ Sequence 268 AA;

Query Match 100.0%; Score 50; DB 19; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRWSS 10  
 Db 259 llmtlrwss 268  
 |||||

RESULT 9

Y69995

XX ID Y69995 standard; Protein; 269 AA.

XX AC Y69995;

XX 31-MAY-2000 (first entry)

XX Human receptor-associated protein from Incyte clone 1361202.

XX Human receptor-associated protein; HRAP; Incyte clone 1361202;  
 KW cytostatic; immunomodulatory; antiinflammatory; cardiant; antianaemic;  
 KW antiarteriosclerotic; hepatotropic; antiarthritic antirheumatic;  
 KW antiasthmatic; osteopathic; antiallergic; antidiabetic; dermatological;  
 KW neuroprotective; diagnosis; treatment; prevention; reproductive disorder;  
 KW cardiovascular; cell proliferative; autoimmune; inflammatory; allergy;  
 KW gastrointestinal; atherosclerosis; cirrhosis; leukaemia; cancer; AIDS;  
 KW arthritis; anaemia; asthma; dermatitis; diabetes; osteoporosis;  
 KW multiple sclerosis; irritable bowel syndrome.

XX Homo sapiens.

XX OS

Key Location/Qualifiers

FT Peptide 1..19

FT /label= Signal\_peptide

FT Protein 20..269

FT /label= Mature\_HRAP

FT Domain 35..111

FT /label= Ig-like\_domain

FT /note= "Signature sequence"

FT Modified-site 19

FT /note= "Potential phosphorylation site"

FT Modified-site 31

FT /note= "Potential phosphorylation site"

FT Modified-site 43

FT /note= "Potential phosphorylation site"

FT Modified-site 72

FT /note= "Potential phosphorylation site"

FT Modified-site 154

FT /note= "Potential phosphorylation site"

FT Modified-site 176

FT /note= "Potential phosphorylation site"

FT Modified-site 221

FT /note= "Potential phosphorylation site"

FT Modified-site 18

FT /note= "Potential phosphorylation site"

FT Modified-site 78

FT /note= "Potential phosphorylation site"

FT Modified-site 89

FT /note= "Potential phosphorylation site"

FT Modified-site 125

FT /note= "Potential phosphorylation site"

FT Modified-site 146

FT /note= "Potential phosphorylation site"

FT Modified-site 173

FT /note= "Potential phosphorylation site"

FT Modified-site 194

FT /note= "Potential phosphorylation site"

FT Modified-site 263

FT /note= "Potential phosphorylation site"

FT Modified-site 41

FT /note= "Potential N-glycosylation site"

FT Modified-site 82

FT /note= "Potential N-glycosylation site"

FT Modified-site 161

FT /note= "Potential N-glycosylation site"

FT Modified-site 195

FT /note= "Potential N-glycosylation site"

FT Modified-site 206

FT /note= "Potential N-glycosylation site"

FT Modified-site 242

FT XX /note= "Potential N-glycosylation site"

PN WO200008155-A2.

XX 17-FEB-2000.

PD 06-AUG-1999; 99WO-US17777.

XX 07-AUG-1998; 98US-0160065.

PR 01-SEP-1998; 98US-0098703.

XX (INCY-) INCYTE PHARM INC.

PA Hillman JL, Yue H, Lal P, Tang YT, Gorgone GA, Guegler KJ;

XX Corley NC, Baughn MR;

PI WPI; 2000-205710/18.

DR N-PSDB; 250897.

XX New human receptor-associated proteins (HRAP) useful for the diagnosis,

XX treatment and prevention of cell proliferative, autoimmune,

XX inflammatory, reproductive, cardiovascular, and gastrointestinal

XX disorders

XX Claim 1; Page 82-83; 99pp; English.

XX The present sequence is a human receptor-associated protein

XX (HRAP) from incyte clone 1361202 obtained from LONGNOT12 cDNA library.

XX This sequence is expressed in hematopoietic/immune, gastrointestinal

XX and reproductive tissues. HRAP has cytostatic, immunomodulatory,

XX antiinflammatory, cardiant, antiarteriosclerotic, hepatotropic,

XX antiarthritic, antirheumatic, osteopathic, antiallergic, antianemic,

XX antiasthmatic, antidiabetic, dermatological and neuroprotective

XX activities. The present sequence is useful in the diagnosis, treatment

XX and prevention of disorders associated with HRAP expression, especially

XX cell proliferative, autoimmune/inflammatory, reproductive,

XX cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,

XX cirrhosis, leukaemia, cancer, AIDS, arthritis, allergies, anaemia,

XX asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and

XX irritable bowel syndrome).

XX Sequence 269 AA;

SQ

Query Match 100.0%; Score 50; DB 21; Length 269;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRWSS 10

DB 260 llmtlrwss 269

RESULT 10

P50256

ID P50256 standard; Protein; 270 AA.

XX AC P50256;

XX DT 07-OCT-1991 (first entry)

XX DE Sequence of T-cell antigen receptor alpha chain encoded by cDNA

XX DE clone T11.

XX Diagnosis; site-directed therapy.

XX KW Homo sapiens.

XX OS

XX Key Location/Qualifiers

XX FT Peptide 1..20

XX FT /label= leader

XX FT 21..114

XX FT Region

XX FT /label= variable

FT Region 115..131

FT /label= joining

FT Region 132..249

FT /label= constant

FT Region 250..265

FT /label= transmembrane

FT Region 266..270

FT /label= cytoplasmic

FT Modified-site 42..44

FT /note= "possible carbohydrate attachment site"

FT Modified-site 200..202

FT /note= "as above"

FT Modified-site 214..216

FT /note= "as above"

FT Modified-site 243..245

FT /note= "as above"

XX WO503947-A.

PN

XX 12-SEP-1985.

XX 28-FEB-1985; 85WO-US00367.

XX 22-OCT-1984; 84US-0663809.

PR 01-MAR-1984; 84US-0585333.

PR 31-OCT-1984; 84US-0666988.

XX (STRD ) LELAND STANFORD UNI.

PA Davis MM, Hedrick SM;

XX WPI; 1985-249152/40.

PI N-PSDB; N50284.

XX New DNA sequences coding for T-cell antigen receptors or

XX fragments - useful in diagnostic assays, affinity chromatography,

XX site directed therapy and diagnosis

XX Disclosure; Fig 3; 41pp; English.

XX Mammalian T-cell receptors appear to be 80-90kdal heterodimers,

XX which are disulphide linked, and composed of two distinct

XX glycoproteins of about 40 to 50kd, referred to as the alpha- and

XX beta- subunits (N50284, P50256; N50280, P50252 respectively). The

XX chains may be divided up into regions associated with specific exons

XX by analogy to immunoglobulins. Genes encoding helper T-cell antigen-

XX specific receptor subunits alpha- and beta- (TH-Ag receptor, alpha-

XX or beta-subunit) were isolated. For the beta-subunit three thymus-

XX derived clones were obtd. designated 86T1, 86T3 and 86T5 (N50280-

XX N50282). The inventors claim a DNA sequence of at least about 15nt

XX present in the sequence of 86T1 (N50280) or T11 (N50284) joined to

XX non-wild type DNA.

XX Sequence 270 AA;

SQ

Query Match 100.0%; Score 50; DB 6; Length 270;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRWSS 10

DB 261 llmtlrwss 270

RESULT 11

W36111

ID W36111 standard; Protein; 272 AA.

XX AC W36111;

XX DT 19-MAY-1998 (first entry)

XX

Mouse T-cell receptor alpha-chain protein from cell line B4-9.52.  
 Mouse; T-cell receptor; alpha-chain constant region; antigen-specific;  
 immunosuppressant; humoral; cell mediated immune response; allergy;  
 hypersensitivity; autoimmune reaction; transplant rejection.

Mus sp.

WO9743411-A1.

20-NOV-1997.

09-MAY-1997; 97WO-JP01565.

29-MAY-1996; 96JP-0135572.

10-MAY-1996; 96JP-0116101.

(KIRI ) KIRIN BEER KK.

Honma N, Mikayama T, Yuyama N;

WPI; 1998-008880/01.

N-PSDB; V01417.

Immunosuppressant peptide containing T-cell receptor alpha-chain  
 sequence - are not antigen-specific and do not induce antibody  
 production

Example 9; Page 39-40; 63pp; Japanese.

The present sequence represents mouse T-cell receptor alpha-chain from  
 cell line B4-9.52. The protein is an immunosuppressant which is not  
 antigen-specific and suppresses both humoral and cell-mediated immune  
 reactions. It can be used for treatment and/or prevention of delayed  
 hypersensitivity reactions, allergies and autoimmune reactions, and  
 inhibition of transplant rejection. The protein does not induce the  
 formation of antibodies against them to any significant extent.

Sequence 272 AA;

Query Match 100.0%; Score 50; DB 19; Length 272;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRLWSS 10

|||||

Db 263 llmtlrlwss 272

RESULT 12

Y05405

ID Y05405 standard; Protein; 273 AA.

AC Y05405;

02-JUL-1999 (first entry)

Killer T-cell receptor protein sequence.

Killer T-cell receptor; HIV; infection; HIV-1 IIIB; therapy.

Mus musculus.

WO9916885-A1.

08-APR-1999.

28-SEP-1998; 98WO-JP04345.

26-SEP-1997; 97JP-0262536.

(KYOW ) KYOWA HAKKO KOGYO KK.

PA (SAIT/) SAITO T.

(TAKA/) TAKAHASHI H.

PI Saito T, Takahashi H;

XX WPI; 1999-255096/21.

DR N-PSDB; X36398.

XX Killer T-cell receptor peptide specifically recognizing HIV-infected

PT cells

XX Example 3; Page 68-70; 75pp; Japanese.

PS This sequence is a killer T-cell receptor protein of the invention,

XX which recognises and damages cells infected with human immunodeficiency

CC virus (HIV), especially with HIV-1 IIIB. The receptor can be used as

CC a constituent of compositions for the treatment of HIV infection.

XX Sequence 273 AA;

SQ

Query Match 100.0%; Score 50; DB 20; Length 273;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLMTLRLWSS 10

|||||

Db 264 llmtlrlwss 273

RESULT 13

Y05404

ID Y05404 standard; Protein; 274 AA.

XX AC Y05404;

02-JUL-1999 (first entry)

Killer T-cell receptor protein sequence.

Killer T-cell receptor; HIV; infection; HIV-1 IIIB; therapy.

Mus musculus.

WO9916885-A1.

08-APR-1999.

28-SEP-1998; 98WO-JP04345.

26-SEP-1997; 97JP-0262536.

(KYOW ) KYOWA HAKKO KOGYO KK.

(SAIT/) SAITO T.

(TAKA/) TAKAHASHI H.

Saito T, Takahashi H;

WPI; 1999-255096/21.

N-PSDB; X36392.

Killer T-cell receptor peptide specifically recognizing HIV-infected

PT cells

XX Claim 7; Page 64-66; 75pp; Japanese.

XX This sequence is a killer T-cell receptor protein of the invention,

CC which recognises and damages cells infected with human immunodeficiency

CC virus (HIV), especially with HIV-1 IIIB. The receptor can be used as

CC a constituent of compositions for the treatment of HIV infection.

XX Sequence 274 AA;

SQ

CC N60079 was generated from the Jurkat human leukaemia T cell line. It is contained in T cell clone pY14. The labelled nucleic acid and monoclonal or polyclonal antibodies to the polypeptides may be used to determine whether unknown cells, e.g. tumour cells, are T cells.

XX  
SQ Sequence 277 AA;

Query Match 100.0%; Score 50; DB 7; Length 277;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LLMTLRWSS 10  
          |||||

DB 268 llmtlrwss 277

RESULT 15  
B63434  
ID B63434 standard; Protein: 99 AA.  
XX  
AC B63434;  
XX  
DT 26-MAR-2001 (first entry)  
XX  
DE Human breast cancer associated antigen protein sequence SEQ ID NO:796.  
XX  
KW Human: breast cancer; gastric cancer; prostate cancer; diagnosis;  
KW cancer associated antigen; cytostatic; cancer vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO2000073801-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 26-MAY-2000; 2000WO-US14749.  
XX  
PR 28-MAY-1999; 99US-0136526.  
PR 10-SEP-1999; 99US-0153454.  
XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
XX  
PI Obata Y;  
XX  
DR WPI; 2001-025274/03.  
XX  
PT Nucleic acids encoding breast, gastric and prostate cancer associated antigen precursors, useful for diagnosing and treating a condition characterized by expression of an abnormal amount of a protein, e.g. cancer -  
XX  
PS Example 1; Page 565; 799pp; English.  
XX  
F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent nucleotide sequences encoding human breast, gastric and prostate cancer associated antigen precursors (CAAP) respectively. B63232 to B63467, B63468 to B63721 and B63722 to B63970 represent human breast, gastric and prostate CAAP protein sequence respectively. CAAPs have cytostatic activity and can be used in the production of cancer vaccines. The human CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are useful for diagnosing and treating a condition characterised by expression of an abnormal amount of a protein, e.g. cancer.

XX  
SQ Sequence 99 AA;

Query Match 84.0%; Score 42; DB 22; Length 99;  
Best Local Similarity 80.0%; Pred. No. 1.1;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LLMTLRWSS 10  
          |||||

Query Match 100.0%; Score 50; DB 20; Length 274;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LLMTLRWSS 10  
          |||||

DB 265 llmtlrwss 274

RESULT 14  
P60065  
ID P60065 standard; Protein: 277 AA.  
XX  
AC P60065;  
XX  
DT 23-JUL-1991 (first entry)  
XX  
DE Sequence of a polypeptide which is at least part of the alpha chain of T cell antigen receptor.  
DE  
XX  
KW Tumour cell; diagnosis; T cell antigen receptor.  
XX  
OS Homo sapiens.  
XX  
FH Location/Qualifiers  
FT Key 1..20  
FT Region /label= leader  
FT Region 21..112  
FT Region /label= variable  
FT Region 113..118  
FT Region /label= diversity  
FT Region 119..135  
FT Region /label= joining  
FT Region 136..252  
FT Region /label= constant  
FT Region 253..272  
FT Region /label= transmembrane  
FT Region 273..277  
FT Region /label= cytoplasmic  
FT Region 42..45  
FT Region /label= potential N-glycosylation site  
FT Region 118..120  
FT Region /label= potential N-glycosylation site  
FT Region 169..171  
FT Region /label= potential N-glycosylation site  
FT Region 203..205  
FT Region /label= potential N-glycosylation site  
FT Region 214..216  
FT Region /label= potential N-glycosylation site  
FT Region 250..252  
FT Region /label= potential N-glycosylation site  
XX  
EP200350-A.  
XX  
XX  
XX 05-NOV-1986.  
XX  
XX 25-MAR-1986; 86EP-0021945.  
XX  
XX 15-APR-1985; 85US-0723306.  
XX  
XX (ONTA-) ONTARIO CANCER INST.  
XX  
XX Mak TW;  
XX  
XX WPI; 1986-292977/45.  
XX  
XX N-PSDB; N60079.  
XX  
XX Nucleic acid encoding T-cell antigen receptor polypeptide - used in identifying T-cells etc.  
XX  
XX Claim 11; Fig 4; 23pp; English.  
XX  
XX



Fri May 11 08:35:06 2001

us-09-202-305-12.rag

Page 9

Db 41 lllxrlwss 50

Search completed: May 10, 2001, 10:10:34  
Job time: 59 sec

